Structural Studies of Fragments of G-Protein Coupled Receptors and their Ligands by NMR

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Alexey Neumoin

aus

Russland

Promotionskomitee

Prof. Dr. John Robinson (Vorsitz)

Prof. Dr. Nathan Luedtke

Prof. Dr. Oliver Zerbe (Leitung der Dissertation)

Zürich, 2009

Table of Contents

Summary

In the course of my doctoral studies I characterized the structure and dynamics of G-protein coupled receptor (GPCRs) fragments and their ligands by high-resolution NMR. The receptors of the GPCR family are transmembrane proteins of prime biological importance. All members of this family possess similar architecture of seven membrane-spanning α-helices and are involved in various signal transduction processes. First part of my work is devoted to the investigation of the structural determinants of the GPCR ligand peptide YY and monitoring the folding process of this peptide in solution. PYY is a 36-residue C-terminally amidated polypeptide that belongs to the neuropeptide Y family of peptide hormones. These molecules are involved in the regulation of a variety of physiological processes, such as for example food uptake. In the second part of my thesis I directed my efforts towards elucidation of the structure and probing the dynamic properties of the transmembrane fragments of the GPCRs in native-like environments. The subject of my studies was the α-factor G-protein coupled Ste2p receptor, which is involved in sensing pheromones in yeast. Two large polypeptide fragments including the first and the second (peptide TM1TM2) and the seventh (peptide TM7) transmembrane domains of the Ste2p receptor were structurally characterized in micellar solution. The obtained results provide important insights into the GPCR architecture in a membrane bilayer.

In the first part of my work I focused on the structural determinants and the folding process of the peptide YY (PYY) in solution. Some of the peptides from neuropeptide Y family adopt a well-defined hairpin structure in water that was first shown for avian pancreatic peptide (aPP) using X-ray crystallography. This helical hairpin is commonly referred to as PP-fold and is characterized by a N-terminal polyproline helix, which is back-folded via a β-turn onto a C-terminal α-helix. The solution structure of the PYY displayed a highly similar helical hairpin, however in the highly homologous neuropeptide Y we were surprised by the absence of the tertiary structure. To investigate the significance of the tertiary contacts, Tyr and Pro residues at the hydrophobic interface of the hairpin-type structure of PYY were

replaced by Ala residues, and the conformational and dynamical properties of the resulting peptides were analyzed by high-resolution NMR spectroscopy. Previously we established the $^{15}N\{^{1}H\}$-NOE as a convenient method to quantify the extent of back-folding. A comparison of the data from different Ala mutant peptides to those of native PYY nicely reflected the differences in backbone rigidity of the N-terminus. Most of the Pro->Ala or the Tyr->Ala mutants possessed increased backbone dynamics, and the differences in N-terminal mobility among them reflected various degrees to which they sample conformations close to the PP-fold. By varying temperature or the methanol content of the aqueous solvent and monitoring chemical shifts we followed the residue-specific formation of tertiary contacts while changing the physical or chemical environment. The PYY peptide in methanol solution was characterized both by determining its solution structure as well as by its internal backbone dynamics as derived from ^{15}N relaxation data. The latter is characterized by a complete loss of tertiary structure. Chemical shifts of Cα in the heat-denaturation experiments displayed sigmoidal curves with very similar points of inflection indicating that both secondary as well as tertiary structure in the heat denaturation was lost synchronously.

The results suggest that helical hairpin formation in PYY peptide is both reversible and cooperative and that specific N- and C-terminal tertiary hydrophobic contacts between the polyproline and the α-helix promote the folding process. In addition, structural analysis of substitutions in the turn region indicates that the loop does not constrain the hairpin structure. The results may also have implications for our understanding of the binding of these peptides to their receptors.

In the second part of the thesis the structure and dynamics of two large fragments of Ste2p the G-protein coupled α-factor receptor from yeast were investigated. Both GPCR fragments were expressed and purified by our colleagues from the group of Prof. Fred Naider (College of Staten Island, NY). At first I investigated the 73-residue (Ste2p(267-339)) peptide TM7 consisting of the 3rd extracellular loop, the 7th transmembrane helix and 40 residues from the cytosolic C-terminal domain in dodecylphosphocholine micelles using solution NMR spectroscopy. The structure revealed the presence of an α-helix in the segment encompassing residues 10 to 30, which was perturbed around the internal Pro24 residue. ^{15}N-relaxation and RDC data supported a rather stable fold for the

transmembrane part of TM7, whereas the exposed segments were more flexible. Spin-label data indicated that the TM7 helix was integrated into dodecylphosphocholine micelles, but displayed flexibility around the internal Pro24 site, exposing residues 22 to 26 to solution and revealed a second site of interaction with the micelle within a region comprising residues 43-58, which formed part of a less well-defined nascent helix. Further I extended my work on a single membrane-spanning TM7 fragment to a longer 80-residue (Ste2p(31-110)) double membrane-spanning peptide TM1TM2, consisting of 19 residues from the N-terminal domain, the 1st transmembrane helix, the first cytoplasmic loop, the second transmembrane helix and 7 residues from the first extracellular loop of the Ste2p receptor. Because of the larger complexity of a double membrane-spanning fragment different isotope labeling patterns were utilized including [^{15}N], [^{15}N, ^{13}C], [^{15}N, ^{13}C, ^{2}H]-labeled and selectively [^{15}N]-labeled at specific amino acid residues or protonated only at selected methyl groups peptides. The structure of TM1TM2 peptide in lyso-palmitoylphosphatidylglycerol micelles revealed the presence of three α-helices encompassing residues 39-47, 49-72 and 80-103, with higher flexibility around the internal Arg58 site of the first transmembrane domain. Several long-range interhelical NOE connectivities supported the folding of TM1TM2 into a tertiary structure forming a crossed helix that splays apart toward the extracellular regions and contains considerable flexibility in the G^{56}VRSG^{60} region. ^{15}N-relaxation and hydrogen-deuterium exchange data support a stable fold for the transmembrane parts of TM1TM2, whereas the solvent-exposed segments were more flexible. Interestingly the NMR structure was consistent with the results of biochemical experiments that identified the ligand binding site within this region of the receptor.

The results obtained during my Ph.D. studies reveal important aspects of the GPCR ligand peptide PYY structure and folding in solution so as shed light on the structure of large fragments of yeast pheromone receptor Ste2p in native-like micellar environment.

Zusammenfassung

Im Laufe meiner Promotion habe ich die Struktur und Dynamik von G-Protein-gekoppelte Rezeptor- (GPCRs) Fragmenten und ihren Liganden mittels hochauflösender NMR charakterisiert. Die Rezeptoren der GPCR-Familie sind Transmembran-Proteine von zentraler biologischer Bedeutung. Alle Mitglieder dieser Familie besitzen eine ähnliche Architektur mit sieben transmembranären α-Helices, und nehmen in verschiedenen Signaltransduktionsprozessen teil. Der erste Teil meiner Arbeit widmet sich der Untersuchung der strukturellen Determinanten des GPCR Liganden Peptid YY und der Verfolgung des Faltungsprozesses dieses Peptids in Lösung. PYY ist ein Polypeptid mit 36 Aminosäuren und C-terminaler Amidierung, das zu der Neuropeptid Y-Familie von Peptid-Hormonen gehört. Diese Moleküle sind in der Regulation einer Vielzahl physiologischer Prozesse involviert, wie zum Beispiel bei der Lebensmittelaufnahme. Im zweiten Teil meiner Arbeit richtete ich meine Bemühungen auf die Aufklärung der Struktur und die dynamischen Eigenschaften der Transmembran-Fragmente der GPCRs in nativen Bedingungen. Das Thema meiner Studien war der α-Faktor G-Protein-gekoppelter Rezeptor Ste2p, der involviert in der Pheromonerkennung in Hefe ist. Zwei große Polypeptid-Fragmente, bestehend aus der ersten und zweiten (Peptid TM1TM2) und der siebten Transmembran-Domän (Peptid TM7) des Ste2p-Rezeptors, wurden in micellärer Lösung strukturell charakterisiert. Die Ergebnisse liefern wichtige Einblicke in die GPCR-Architektur in einem Membran-Bilayer.

Im ersten Teil meiner Arbeit konzentrierte ich mich auf die strukturellen Faktoren und den Faltungsprozess des Peptid YY (PYY) in Lösung. Einige der Peptide aus Neuropeptid Y-Familie haben eine klar definierte hairpin-Struktur in Wasser; diese wurde zum ersten Mal gezeigt für das Avian Pankreas-Peptid mittels Röntgenstrahl-Kristallographie. Dieser helikale ‘hairpin’ wird gemeinhin als PP-fold bezeichnet und besteht aus einer N-terminalen Polyprolin-Helix, die zurückfaltet über einen β-turn auf eine C-terminale α-Helix. Die Lösungsstruktur des PYY zeigt einen sehr ähnlichen helikalen ‘hairpin’, jedoch im hoch-homologen Neuropeptid Y beobachteten wir zu unserem Erstauenen keine Tertiärstruktur. Um die Bedeutung der tertiären Kontakte zu untersuchen, wurden Tyr- und Pro-Reste an der hydrophoben

Oberfläche der ‘hairpin’-Struktur von PYY ersetzt durch Alanin und die konformationellen und dynamischen Eigenschaften der resultierenden Peptide wurden analysiert mittels hochauflösender NMR-Spektroskopie. Zuvor haben wir die $^{15}N\{^{1}H\}$-NOE als eine passende Methode zur Quantifizierung des Umfangs der Rückfaltung etabliert. Ein Vergleich der Daten aus unterschiedlichen Ala-Peptid-Mutanten mit dem nativen PYY spiegelt schön die Unterschiede in der Steifheit des ‘backbones’ des N-Terminus wieder. Die meisten der Pro-> Ala oder der Tyr-> Ala Mutanten besaßen eine erhöhte ‘backbone’-Dynamik, und die Unterschiede in der N-terminalen Mobilität unter ihnen spiegelt verschiedene Grade wieder, zu dem sie Probe Konformationen annimmt, die dem ‘PP-fold’ ähneln. Durch Variation der Temperatur oder des Methanolgehalts des wässrigen Lösungsmittels und Verfolgung des ‘chemical shift’ konnten wir die aminosäure-spezifische Bildung der Tertiärkontakte während der Änderung der physikalischen oder chemischen Umgebung verfolgen. Das PYY Peptid in Methanollösung wurde charakterisiert sowohl durch die Bestimmung seiner Lösungsstruktur als auch durch ihre interne ‘backbone’-Dynamik mittels ^{15}N-relaxation-Daten. Die ‘backbone’-Dynamik zeichnet sich durch einen vollständigen Verlust der tertiären Struktur aus. Die ‘Chemical shifts’ der Cα in den Hitze-Denaturierungs-Experimenten zeigten sigmoidale Kurven mit sehr ähnliche Wendepunkten, was darauf hinweist, dass sowohl Sekundär- als auch Tertiärstruktur in der Hitzedenaturierung synchron verloren werden.

Die Ergebnisse deuten darauf hin, dass die Bildung des helikalen ‘hairpin’ im PYY Peptid reversibel und kooperativ ist und dass spezifische N-und C-terminale hydrophobe Tertiärkontakte zwischen der Polyprolinhelix und der α-Helix den Faltungsprozess fördern. Darüber hinaus deutet die Strukturanalyse von Substitutionen in der ‘turn’-Region darauf hin, dass der ‘loop’ die ‘hairpin’-Struktur nicht hemmt. Die Ergebnisse können auch Auswirkungen für unser Verständnis der Bindung dieser Peptide auf ihren Rezeptoren haben.

Im zweiten Teil der Dissertation wurde die Struktur und Dynamik von zwei großen Fragmenten von Ste2p, dem G-Protein-gekoppelten α-Faktor-Rezeptor von Hefe untersucht. Beide GPCR-Fragmente wurden exprimiert und aufgereinigt von unseren Kollegen aus der Arbeitsgruppe von Prof. Fred Naider (College of Staten Island, NY). Zuerst untersuchte ich das 73-aminosäure-Peptid TM7 (Ste2p (267-339)) bestehend aus dem dritten extrazellulären ‘loop’, der siebten Transmembran-Helix

und 40 Aminosäuren aus der zytosolische C-terminalen Domäne in Dodecylphosphocholin-Micellen mittels NMR-Spektroskopie. Die Struktur offenbarte die Anwesenheit einer α-Helix im Segment von Aminosäurerest 10 bis 30, die um das interne Pro24 gestört wird. ^{15}N-relaxation und RDC-Daten unterstützten einen recht stabilen 'fold' für den Transmembran-Anteil des TM7, hingegen die ausgesetzten Segmente waren flexibler. Die Spin-Label-Daten weisten darauf hin, dass die TM7-Helix in die Dodecylphosphocholin-Micellen integriert wurde, aber zeigten Flexibilität rund um das interne Pro24, da die Aminosäuren 22 bis 26 in die Lösung zeigen, desweiteren zeigten sie einen zweiten Interaktionsort mit der Micelle innerhalb der Region von Aminosäurerest 43 bis 58, die einen Teil einer weniger gut definierten im Entstehen begriffenen Helix bildet. Im weiteren verlängerte ich meine Arbeit an einem einfachen Transmembran-Fragment TM7 zu einem längeren 80-Aminosäure-Doppel-Transmembran-Peptid TM1TM2 (Ste2p (31-110)), bestehend vom 19 Aminosäuren aus der N-terminalen Domäne, die erste Transmembran-Helix, der erste zytoplasmatische 'loop', die zweite Transmembran-Helix und 7 Aminosäuren aus dem ersten extrazellulären 'loop' des Ste2p-Rezeptors. Aufgrund der größeren Komplexität des doppelten Transmembran-Fragments wurden verschiedene Isotopen-Labeling-Muster genutzt: [^{15}N], [^{15}N, ^{13}C], [^{15}N, ^{13}C, ^{2}H]-markiert und selektiv [^{15}N]-markiert an bestimmten Aminosäuren oder protoniert nur an ausgewählten Methyl-Gruppen-Peptiden. Die Struktur des TM1TM2-Peptids in LYSO-palmitoylphosphatidylglycerol-Micellen zeigte das Vorhandensein von drei α-Helices, von Aminosäure 39-47, 49-72 und 80-103, mit einer größeren Flexibilität rund um das interne Arg58 der ersten Transmembran-Domäne. Mehrere 'long range-interhelical NOE' Verbindungen unterstützen die Faltung von TM1TM2 in eine Tertiärstruktur, die eine gekreuzte Helix bildet, die sich ausdehnt in Richtung der extrazellulären Regionen und die erhebliche Flexibilität in der G^{56}VRSG^{60} Region enthält. ^{15}N-relaxation- und Wasserstoff-Deuterium-Austausch-Daten unterstützten einen stabilen 'fold' für die Transmembran-Teile von TM1TM2, während die lösungsmittel-exponierten Segmente flexibler waren. Interessanterweise ist die NMR-Struktur im Einklang mit den Ergebnissen der biochemischen Experimente, die die Ligandenbindungsort in dieser Region des Rezeptors identifizierten.

Die erzielten Ergebnisse während meiner Promotionsstudien zeigen wichtige Aspekte der GPCR-Peptid-Liganden PYY-Struktur und seiner Faltung in der Lösung,

sowie geben sie Aufschluss über die Struktur der großen Fragmente des Hefe-Pheromon-Rezeptor Ste2p in nativer Micellenumgebung.

Abbreviations

GPCR	G-protein coupled receptor
DPC	dodecylphosphocholine
HPLC	high performance liquid chromatography
$^{3}J_{HN\alpha}$	vicinal spin-spin coupling constant between the backbone amide proton and the α proton
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
HSQC	heteronuclear single quantum coherence
$^{15}N\{^{1}H\}$-NOE	heteronuclear Overhauser enhancement of ^{15}N after saturation of ^{1}H
RMSD	root-mean-square deviation
TOCSY	total correlation spectroscopy
2D / 3D	two-dimensional / three-dimensional
CD	circular dichroism
LPPG	1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)]
DOTA	1,4,7,10-tetraazocyclododecane-N,N-acetic acid
TM	transmembrane
IMP	integral membrane protein
Doxyl	(4,4-dimethyl-3-oxazolidine-N-oxyl)
PP	pancreatic polypeptide
PYY	peptide YY (h, human; p, porcine)
NPY	neuropeptide Y
R1	longitudinal relaxation rate constant
R2	transverse relaxation rate constant

1. Introduction

Cell – cell communication is a vital event involved in many biological processes such as cell migration, organ formation, immune response and microbial infection. It forms a solid barrier on the way of the information transfer from one cell to another. The cell membrane keeps the balance between the interior of the cell and the outside world allowing active transport and signaling utilizing diverse biomolecules incorporated in the membrane bilayer. The separate compartments (organelles) inside the cell are surrounded by their own membranes. Cellular membranes have diverse functions in the different regions and organelles of a cell. However, at the microscopic level, they share the common structure of a lipid bilayer. Different biomolecules can be integrated into or bound to the surface of membranes. These components of the membrane introduce wide range of functions – cell protection and insulation, signal transduction and active transport into and out of the cell, etc. Cell signaling processes in many cases are mediated by interactions of membrane proteins and their ligands – relatively small biomolecular compounds (e.g. peptides). A huge class of membrane proteins named G-protein coupled receptors is involved in a large variety of signaling processes. The determination of the structure of G protein-coupled receptors remains an unfulfilled goal of structural biologists with the two exceptions of high-resolution structures of rhodopsin and the β2-adrenergic receptor solved by X-ray crystallography (Palczewski, Kumasaka et al. 2000; Cherezov, Rosenbaum et al. 2007; Rasmussen, Choi et al. 2007). These membrane-bound, information-transducing proteins comprise a major family of proteins with thousands of representatives. Moreover, the impressive successes in the development of pharmaceuticals targeting G protein-coupled receptors (GPCRs) have created much interest in their structure and mechanism of action. Besides the determination of the GPCRs structure the important question remains regarding their mechanism of activation. For many of these receptors the ligands are small water-soluble peptides. These molecules are much better objects for structural studies, especially for NMR, due to their good solubility in water and their much lower molecular weight. However, the solution structure of these peptides often may not correspond to the structure in the receptor-bound form. Therefore, a methodology for

investigation of the structural transition of the peptide from bulk solution to the membrane and the receptor active pocket is of high interest.

This work is devoted to investigation of structural determinants in neuropeptides as the signal-triggering ligands for the cell communication process and to the study of membrane-embedded G-protein coupled receptors, as the receivers and transducers of the information signal. The aim of this study is to improve our understanding of the underlying fundamental processes during ligand recognition by the receptor, and therefore involves studies of folding of the ligands or the structure of the GPCR themselves.

1.1. Membrane biophysics

Structure of lipid bilayer

Biological membranes are important for their role in the maintenance of the cell integrity and for acting as barriers regulating the movement of molecules between cells and between sub-compartments within the cell. Communication through these barriers requires either some kind of signal transduction across the membrane or the passage of molecules through the membrane involving specific transporters. Both of these processes invariably involve a protein or protein assemblies, which require a specific lipid environment for their functional integrity. The subtle regulation of the transducer/transporter is affected through the lipid environment. Membranes are complex structures and for biophysical studies simplifying models are required. Biophysical studies can give very detailed information on the involved molecules, e.g. about their structure and dynamics as well as on the interaction within the membrane, which are generally not accessible from studies of whole cells. To evaluate the results obtained using membrane models, knowledge about the properties of biological and model membranes, their differences and limitation, is of prime importance.

Biological membranes consist of different lipids. Lipids are amphiphilic molecules containing hydrophobic tails and hydrophilic head groups. In aqueous solution they are oriented such that the headgroups are exposed to the water and the

hydrophobic chains are shielded from solvent access. Most membrane phospholipids are cylindrically shaped and therefore form bilayers upon association (Fig. 1B). As shown in X-ray studies many biological membranes are about 5 nm thick (Walker, Carne et al. 1979). Across the bilayer there is a steep gradient of polarity (White and Wimley 1998). In the region of the polar lipid headgroups, the charge density is high, whereas towards the centre of the membrane in the hydrocarbon core the polarity is low.

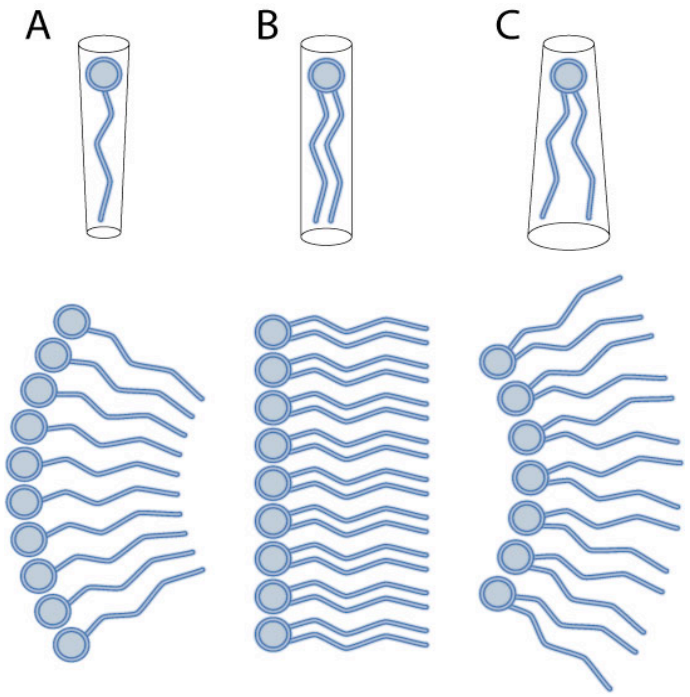


Figure 1 Different curvature of membranes depending on the form of a lipid molecule.

The today well-accepted fluid mosaic model of lipid bilayer membranes, established by Singer and Nicolson in 1972 (Singer and Nicolson 1972), describes the essential features of the biological membrane. It is a two-dimensional fluid, or liquid crystal, in which the hydrophobic integral components such as lipids and membrane proteins are constrained within the plane of the membrane. Lipids and proteins can freely diffuse laterally in the membrane with a diffusion rate of approximately 10^{-8} cm^2/sec (Kornberg and McConnel 1971) and rapid rotation about their long axis can occur. In contrast, transverse diffusion (flip-flop) is very rare, only once every several hours (Devaux 1993). The membrane fluidity depends on the lipid composition and

also largely on the temperature. Bacteria regulate fluidity by varying the degree of unsaturation and the length of fatty acid chains in lipids. Many (eukaryotic) cell membranes contain cholesterol and glycolipids. Cholesterol orients in the bilayer with its hydroxyl group close to the polar headgroups of the phospholipid molecules. The rigid steroid skeleton interacts with the first few CH_2 groups of the phospholipid hydrocarbon chains, thereby making the membrane less fluid and decreasing the permeability of the bilayer to small water-soluble molecules (Spector and Yorek 1985; Bastiaanse, Hold et al. 1997). Regions in which certain molecules such as cholesterol are clustered, called lipid rafts, and in which some membrane proteins tend to accumulate, have become a focus of membrane biology (Munro 2003; Parton and Richards 2003). The major components of membranes comprise lipids, proteins, glycolipids and glycoproteins. The function of the glycolipids is not fully understood, they are only found in the extracellular leaflet of the bilayer as a result from glycosylation in the lumen of the Golgi apparatus. A variety of lipids are found in biological membranes with different chemical nature of headgroups, chain lengths and degree of saturation. In general, lipids are fatty acids and their naturally occurring derivatives (esters or amides). A broader definition includes also compounds related closely to fatty acid derivatives through biosynthetic pathways (e.g. prostanoids, aliphatic ethers or alcohols) or by their biochemical or functional properties (e.g. cholesterol). Important classes are glycerophospholipids, sphingolipids and sterols. Lipids are either zwitterionic (no net-charge) or negatively charged. The lipid composition is tissue dependent and can be different for the inner and outer leaflet (Gennis 1989). Despite its biochemical complexity a simplified treatment of the biomembrane is useful and sufficient for many aspects. The membrane can be subdivided into two distinct phases/compartments: the interfacial (or fixed charge) and the hydrophobic (or hydrocarbon core) compartment. The interfacial region is mainly built by the phospholipid headgroups and their hydrating water molecules (Fig. 2A). It presents opportunities for dipole-dipole interactions and promotes hydrogen bonding with appropriate aminoacid side-chains. The interfacial environment is not a sharp cut-off between aqueous and apolar domains. It rather can be thought of as a gradual transition spanning ca 15Å from the bulk aqueous phase to an anhydrous

hydrocarbon phase (Wiener and White 1992; Sanders and Schwonek 1993; White and Wimley 1999) (Fig. 2A,B). In the transition region the effective dielectric constant gradually decreases, but there are still polar moieties present from the lipid head group and glyceride regions along with a significant concentration of water (Luca, Heise et al. 2003) (Fig. 2B).

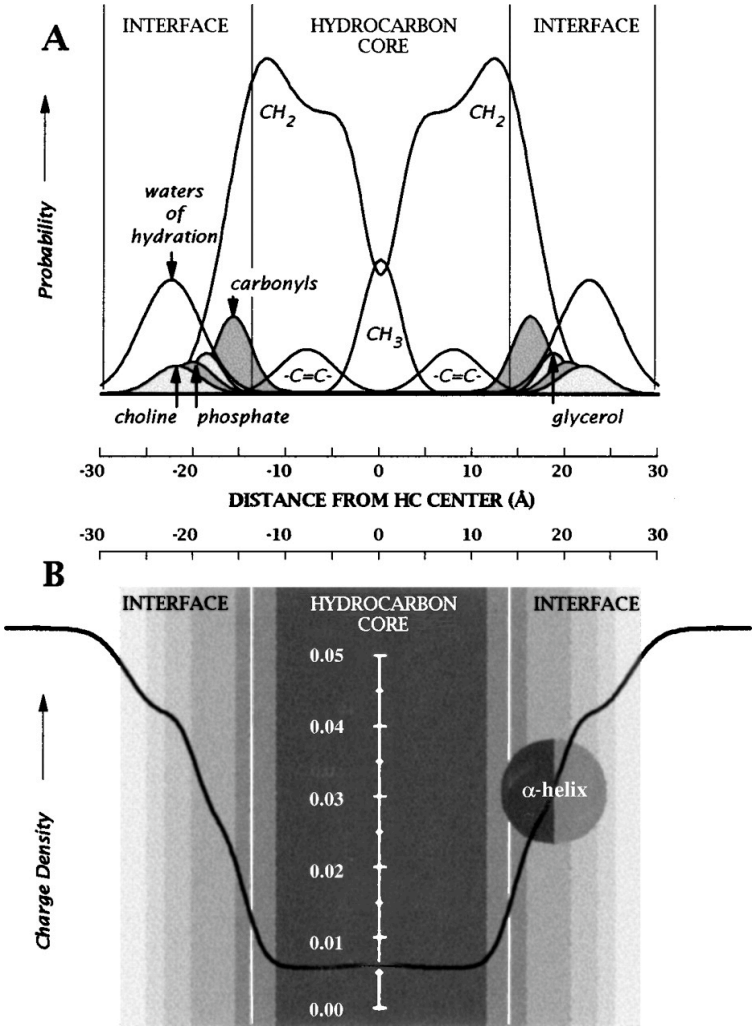


Figure 2 The structure of a fluid liquid-crystalline bilayer and its polarity profile: (A) The structure of a dioleoylphosphocholine (DOPC) bilayer; (B) Polarity profile derived from the group distributions of panel A (White and Wimley 1999).

The hydrophobic core of the lipid bilayer comprises the hydrocarbon chains of the phospholipids. Typical glycerophospholipids can be described as being roughly cylindrical in shape with a cross-section (circular) surface area in the range of 65Å^2 (diameter of ca 9Å; (Lewis and Engelman 1983). The hydrophobic span across a biological membrane is in the range of 30-50Å. A native biological membrane may

have hundreds of distinct lipids (Dowhan 1997) mostly of phospholipids or glycolipids. Lipids in native membranes are highly dynamic and randomly laterally diffuse around the more immobile embedded proteins. Membrane proteins usually rotate around the bilayer normal within native-like membranes, with correlation times for this axial rotation typically being in the microsecond range (Park, Mrse et al. 2006).

Membrane model systems

Native membranes

There are published examples of NMR studies using samples involving non-purified membrane proteins in native membranes. One set of studies involves titrating membranes rich in the protein of interest with ligands specific to that protein. The NMR spectrum of the ligand is monitored. For solid state NMR, the ligand is generally labeled with ^{13}C or some other observable heteroatom (Watts, Ulrich et al. 1995). Solution NMR studies of the conformation of membrane protein-associated ligand may be possible in cases where exchange between free (isotropic) ligand and receptor-bound ligand is rapid on the NMR time scale. In such cases, transferred nuclear Overhauser effect (NOE) measurements may be possible (Kisselev, Kao et al. 1998; Koenig, Mitchell et al. 2000). A second class of studies involves detection of the membrane protein itself. This generally requires a biosynthetic isotopic labeling method. Along with the protein of interest many other proteins in the membrane may also be labeled. In some cases, this is not a serious problem because the protein of interest is in vast excess to other proteins within a given membrane preparation as a result of natural or induced overexpression (Griffiths, Lakshmi et al. 1994; Smith, Aschheim et al. 1996). In other cases, double isotopic labeling schemes and accompanying pulse technology can be employed to filter out unwanted signals and to focus upon specific protein sites of interest (Seigneuret and Kainosho 1993).

Vesicles

Vesicles are spherical compartments of bilayered/multilayered phospholipids enclosing a liquid (usually water) (Fig. 3C). The variables distinguishing vesicle types from each other are degree of bilayer multilayering, size, lipid composition, net

charge, shape and lyotropic phase. A vast amount of literature describes the preparation and characterization of lipid vesicles (liposomes) and reconstituting membrane proteins into them (Rand and Parsegian 1989; Bloom, Evans et al. 1991; Silvius 1992; Rigaud, Pitard et al. 1995; Rigaud, Levy et al. 1998). Because even the smallest vesicles are very large from a solution NMR point of view (MDa aggregate weights), vesicles are not used in direct solution NMR studies of integral membrane proteins (Henry and Sykes 1994). However, they are broadly employed in solid state NMR and other spectroscopic methods as CD spectroscopy and fluorescence spectroscopy. In most cases, large multilamellar vesicles are used because these are easily formed and because multilayering makes it possible to prepare samples with relatively high concentrations of lipid and membrane proteins, a fact, which is critical for optimizing NMR signal intensity. It may be particularly important to choose bilayers in which the transbilayer thickness matches the span of the hydrophobic domain of the membrane protein of interest (Killian 1998).

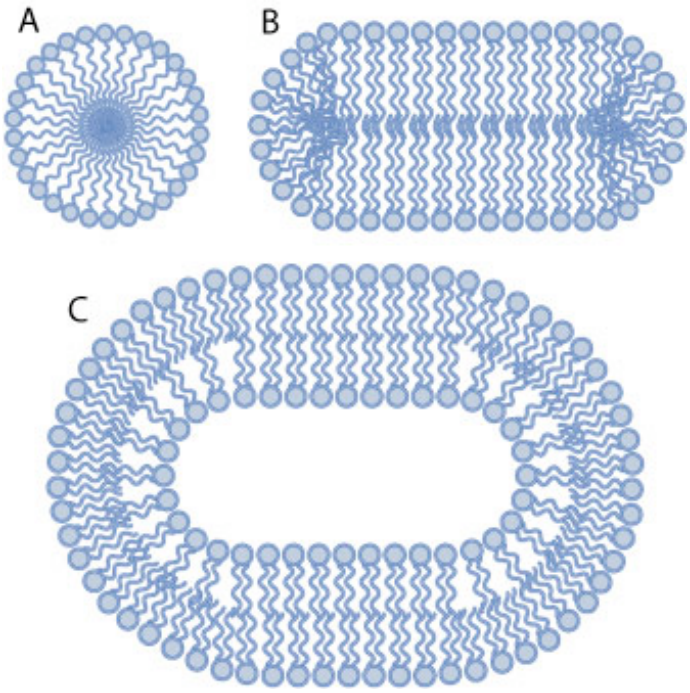


Figure 3 Different bilayer model systems: A - micelle, B - bicelle, C – vesicle.

Bicelles

Bicelles represent an intermediate case between lipid vesicles and classical mixed micelles, being composed of long-chain phospholipids that form the flat surface and short-chain phospholipids that make up the rim (Ram and Prestegard 1988). Unlike lipid vesicles, bicelles do not have inner aqueous compartments and are optically clear; unlike classical mixed micelles, they retain a bilayered domain, which maintains a number of key dynamic and conformational properties of liquid crystalline phase bilayers (Fig. 3B). A large number of solution NMR structural studies have been conducted on small membrane peptides using fast tumbling bicelles rather than detergent micelles (Vold, Prosser et al. 1997). Larger bicelles can be oriented by magnetic fields such that their bilayer normals are orthogonal to the direction of the applied field (Sanders, Hare et al. 1994). The best-characterized bicellar systems are composed of mixtures of dimyristoylphosphatidylcholine (DMPC) and dihexanoylphosphatidylcholine (DHPC) as the short-chain component. For these mixtures, bicelles form over a fairly wide DMPC:DHPC ratio and over a fairly wide range of temperatures, but only above the phase transition temperature of the lipid component. There are known successful examples of membrane proteins incorporation into bicelles in a manner consistent with maintenance of the bicellar morphology (Sanders and Landis 1995). Bicelles are useful for solid-state NMR studies of membrane proteins and extremely beneficial in order to determine the residual dipolar couplings of the water-soluble proteins (Tjandra and Bax 1997; Delaglio, Kontaxis et al. 2000).

Micelles and mixed micelles

Detergent micelles are useful as a medium in which to solubilize membrane proteins for solution NMR because of their relatively small size (usually 10–100 kDa) in comparison to any available bilayer assemblies. Micelles form spontaneously when detergents with a headgroup larger than the aliphatic chain are dissolved (Fig. 1A, Fig. 3A). There is vast number of different detergent types of varying charges and distinctly varying molecular topologies (Lichtenberg, Robson et al. 1983; Neugebauer 1988). For NMR experiments, the primary classes employed to date are those having a single polar head group and a single extended apolar tail type, such as dodecylphosphocholine (DPC) or sodiumdodecylsulfate (SDS). Recently lyso-

glycerophospholipids appear to be a convenient media for solubilizing and NMR measurements of integral membrane proteins (Krueger-Koplin, Sorgen et al. 2004; Sanders and Sonnichsen 2006). Detergent micelles are often thought of as spherical assemblies. However, in reality most micelles are somewhat cigar-shaped (prolate ellipsoid) or discoidal (oblate ellipsoid) (Tanford 1980; Herzfeld 1996). It should also be noticed that when lipids or proteins are added to micelles, both the critical micelle concentration (CMC) and the aggregation number can be perturbed dramatically (Moller and le Maire 1993; Hink, van Hoek et al. 1999). This is especially true for membrane proteins when the size of the protein approaches or exceeds the normal size of the protein-free micelle. In these cases, the properties of the protein primarily determine the final detergent-protein aggregate size. The term "mixed micelles" usually implies a lipid component. For micelle size to remain small enough for solution NMR studies, the lipid/detergent ratio must remain low. The usual reason for having lipid present in studies of membrane proteins involving detergent micelles is to enhance membrane protein stability and functionality.

Organic solvent mixtures

Complex membrane proteins can often be solubilized in organic solvent mixtures. The concept of "naked" membrane proteins in such mixtures is appealing for NMR because the effective MW of the protein is not increased by the association of detergents or other amphiphiles. Girvin and co-workers have shown that subunit c of the F1-F0 ATPase (two transmembrane helices with a short loop connector) adopts a stable, native-like fold in at least one organic solvent system and have determined its structure at high resolution (Girvin, Rastogi et al. 1998; Rastogi and Girvin 1999). While the secondary structure may often be retained, it highly probable that significant perturbations of protein tertiary structure will usually accompany solubilization by this method, as it was shown for the membrane protein from *E.coli* diacyldlycerol kinase (Vinogradova, Badola et al. 1997) and recently for yeast pheromone receptor Ste2p transmembrane fragments studied in organic solvents and solution micellar conditions (Arshava, Taran et al. 2002; Estephan, Englander et al. 2005; Neumoin, Arshava et al. 2007).

Membrane proteins

Classification

There are two main classes of membrane proteins: integral (intrinsic) and peripheral (extrinsic). Peripheral membrane proteins are mostly water-soluble proteins attached to the membrane with an anchoring group. Often these proteins can be studied in the soluble form by clipping off the anchor or by finding conditions in which the anchor dissociates from the membrane and also becomes water-soluble. Integral membrane proteins in contrast have (multiple) transmembrane segments that are of particular interest. Roughly 20 – 35% of all proteins are integral membrane proteins and until now from thousands of protein structures that have been solved at high resolution, only about 140 represent membrane proteins (for information see S. H. White group web-site: http://blanco.biomol.uci.edu/MemPro_resources.html). The vast majority of the membrane proteins are α-helical, only a small class of proteins is formed by the β-barrel proteins that consist of β-sheets. The α-helical proteins possess one or more highly hydrophobic stretches of about 20-25 residues, each of which span the native lipid bilayer as an α-helix. The helical structure is stable in the hydrophobic environment of the bilayer interior because it completely satisfies the hydrogen bonding requirements of the polypeptide backbone (Henry and Sykes 1994). In a protein with a single TM helical domain the TM helix may simply act as a membrane anchor, to localize protein and hence its biochemical activity in proximity of the membrane. However, if a polypeptide chain possesses many transmembrane helical spans (polytopic protein) or forms an oligomeric complex with other transmembrane α-helices, then the bundle of helices within the bilayer is closely involved in the function of the protein or protein complex. Important examples for polytopic helical proteins are the G-protein-coupled receptors (see below) and ion channels. Membrane proteins built mainly of β-sheet structures are found in the outer membrane of gram-negative bacteria and in mitochondria. They exhibit a completely different strategy for satisfying the hydrogen bonding capacity of the polypeptide backbone in an apolar medium and possess no continuous hydrophobic sequences but hydrophobic residues form the outer surface of a membrane-spanning β-barrel (Cowan, Schirmer et al. 1992). Unified by many common structural features, β-barrel membrane proteins carry out diverse functions in different organisms (Wimley 2003).

Structure prediction is more difficult than for α-helical proteins, but several algorithms have been developed to overcome existing difficulties (Wimley 2002).

The membrane is more highly organized, than an isotropic liquid and this organization might help to stabilize membrane proteins. The lipid-facing regions of the transmembrane domain of membrane proteins are formed mostly by the hydrophobic aminoacids. Within the transmembrane domain side chains that do not face the lipid may either interact at a protein interface or line an aqueous pore or cavity. In the latter case residues are mostly polar. However, protein–protein interfaces within the transmembrane domains are typically apolar, just as for the interior of water-soluble proteins (Tian, Karra et al. 2005). The hydrogen-bonding propensity of backbone amide moieties within transmembrane segments is almost fully maximized.

Transmembrane α-helices architecture

Membrane-spanning helices found in membrane proteins can be divided into two regions: a hydrophobic transmembrane (TM) segment, which interacts with the hydrophobic core of the bilayer, and the residues flanking the membrane-spanning segments that are located at the membrane-water interfacial region (White and Wimley 1999). The hydrophobic core of the lipid bilayer is ~30 Å thick, while the membrane-water interfacial region makes up an ~15 Å thick region on both sides of the membrane (White and Wimley 1999). The length of an α-helix increases ~1.5 Å per residue, therefore membrane-spanning α-helices have to be ~20 residues long or more to traverse the hydrophobic core of the lipid bilayers.

On the basis of known membrane protein structures, the hydrophobic segment of TM helices has been reported to be 20-25 residues long (Wallin and von Heijne 1998). These protein segments are highly enriched in hydrophobic residues like valine, leucine, alanine, isoleucine, and phenylalanine (Arkin and Brunger 1998; Ulmschneider and Sansom 2001). Most of these residues are distributed along the TM segments, but valine and leucine are found more frequently in the center of the bilayer (Arkin and Brunger 1998). In multispanning proteins, alanine has a preference for the interior of α-helical bundles while bulky residues tend to be on the "outer side" (Ulmschneider and Sansom 2001). Polar residues are also an important part of TM helices. Since such residues tend to be shielded from the hydrophobic acyl chains,

they are likely to play a role in helix-helix association and hence in the folding of multispanning membrane proteins. The most frequently found polar residues in TM segments are threonine and serine (Arkin and Brunger 1998). These residues seem to have no preference for either transmembrane or non-transmembrane segments (Ulmschneider and Sansom 2001). Their distribution in membrane proteins may be explained by their ability to form hydrogen bonds with adjacent carbonyl oxygens (Gray and Matthews 1984). Ionizable residues are found even less frequently in TM segments than polar residues. They may play important roles in membrane protein function, for example, in bacteriorhodopsin, where ionizable residues are involved in proton transfer (Henderson, Baldwin et al. 1990). Glycine is fairly common in TM -helices, where it may facilitate closer packing of the helices. This is probably why glycine is found more frequently in the interior of multispanning proteins than in single helices (Eilers, Shekar et al. 2000; Ulmschneider and Sansom 2001). Proline is found only to a small extent within TM helices (Ulmschneider and Sansom 2001), where it seems to be preferentially localized to the center of the bilayer (Cordes, Bright et al. 2002). It induces kinks or bends and may have an important role in membrane protein folding. Proline also may form molecular hinges that function as conformational switches in the TM helices of channels and receptors (Lu, Marti et al. 2001). Compared to the hydrophobic acyl chain region of the bilayer, the interfacial region presents a complex environment with its carbonyl groups, lipid headgroups, and water molecules. Analysis of known structures of membrane proteins has shown that the interfacial regions are enriched especially in the aromatic residues tryptophan and tyrosine and the hetero-aromatic residue histidine (Wallin, Tsukihara et al. 1997; Ulmschneider and Sansom 2001). An increased affinity of these residues for the interfacial region is mostly because their hydrogen bond forming ability and their dipolar but still hydrophobic character (Yau, Wimley et al. 1998; de Planque, Bonev et al. 2003). Indeed, phenylalanine, which is aromatic but hydrophobic and lacking hydrogen bond forming ability, is distributed preferentially to the hydrophobic TM segment (Wallin, Tsukihara et al. 1997; Ulmschneider and Sansom 2001). The charged residues lysine, arginine, aspartate, and glutamate are frequently found more outward at the flanks of TM helices (Ulmschneider and Sansom 2001). In contrast to tryptophans, these residues play an important role in determining the topology of membrane proteins and are distributed according to the “positive-inside” rule. That is,

when newly synthesized membrane proteins insert into a membrane, the positively charged lysine and arginine residues preferentially will stay at the "cis" side, while the negatively charged aspartate and glutamate residues are found more frequently at the "trans" side of the membrane (Wallin, Tsukihara et al. 1997). Nonpolar residues in TM segments tend to point toward the middle of the bilayer, while polar residues have a tendency to stretch toward the aqueous phase (Chamberlain, Lee et al. 2004; Granseth, von Heijne et al. 2005). In particular, the longer, positively charged side chains in fact may reach the interfacial region from a position deeper in the hydrophobic part of the bilayer by what is known as snorkeling. This allows more flexibility in the positioning of side chains at the interface and hence more freedom in the localization and dynamics of the helices in the bilayer (Strandberg and Killian 2003).

Membrane proteins affect lipids organization

Proteins might affect lipids in different ways, as illustrated in Fig. 4. Much understanding has been obtained from studies of the consequences of hydrophobic mismatch in model systems. More than 20 years ago, Mouritsen and Bloom (Mouritsen and Bloom 1984) suggested in their "mattress model" that lipids adapt to mismatch by stretching or disordering, as indeed observed for model peptides by simulations (Kandasamy and Larson 2006) (Fig. 4A). Also, experimental ^{2}H NMR studies on WALP peptides showed a systematic response of lipid chain order to mismatch (de Planque, Greathouse et al. 1998). However, using X-ray diffraction of similar samples, Weiss and co-workers (Weiss, van der Wel et al. 2003) could not detect any bilayer thickness adaptation, most likely because the adaptations were too small. However for the β-helical peptide gramicidin the X-ray measurements did detect an increase in bilayer thickness (Weiss, van der Wel et al. 2003). This may mean that proteins with larger cross-sectional areas cause more adaptation of the surrounding lipids.

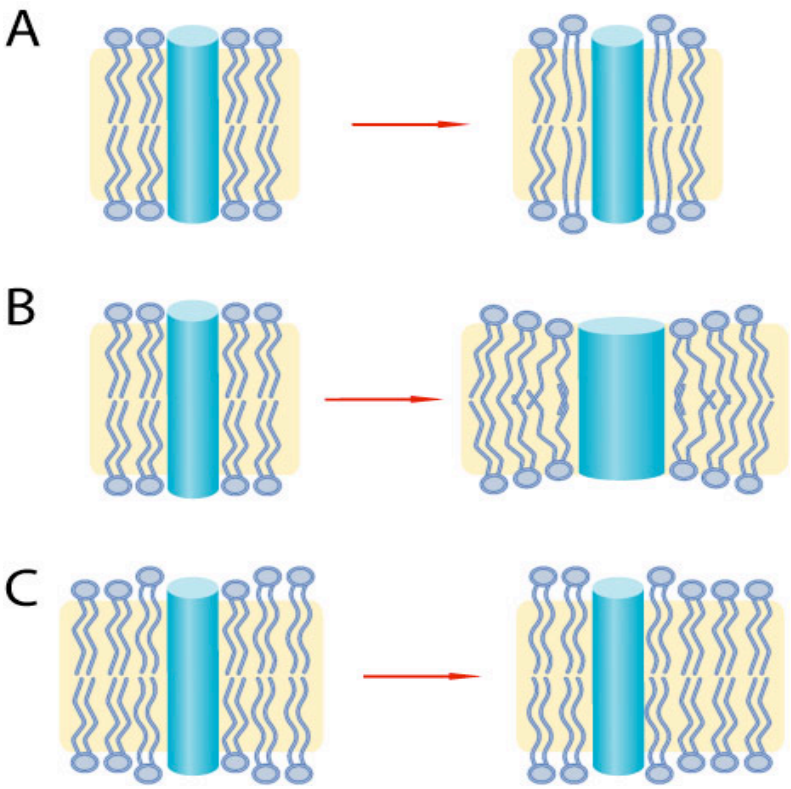


Figure 4 Different mechanisms showing how protein affects lipids organization (see text for details).

For a variety of lipids, incorporation of short peptides was found to promote the formation of phases with negative curvature (Siegel, Cherezov et al. 2006) (Fig. 4B). This can be understood in terms of an adaptation of the lipids by disordering their chains. Similarly, incorporation of relatively long peptides may promote the formation of structures with positive curvature (Fig. 4C). A combination of positive and negative mismatch may serve as a mechanism in nature to stabilize the formation of highly curved membranes in processes such as fusion or fission. In biological membranes, mismatch also could play a role in lateral organization of the lipids, if proteins attract lipids with the best-matching hydrophobic length. However, studies by different groups on model peptides in binary lipid mixtures in the fluid phase did not show a significant degree of such sorting (Ridder, Spelbrink et al. 2004). Also, in this case, it was suggested that the size or cross-sectional diameter of the protein may be an important factor (Ridder, Spelbrink et al. 2004).

Membrane proteins adapt to lipid bilayers

General properties of lipids that can influence membrane proteins include lipid packing, the lateral pressure profile across the membrane, intrinsic curvature of the lipids, bilayer thickness and electrostatic properties (Lee 2004; McIntosh and Simon 2006). In particular, studies of the consequences of varying bilayer thickness have been useful as a tool to determine basic principles of protein–lipid interactions (de Planque and Killian 2003). When the hydrophobic thickness of the bilayer is larger or smaller than the hydrophobic length of the transmembrane protein, this would lead to exposure of hydrophobic groups to a hydrophilic environment. As a consequence, proteins and lipids will tend to adapt their organization, for example in the case of positive mismatch, that is, when the transmembrane segments are relatively long (Fig. 5).

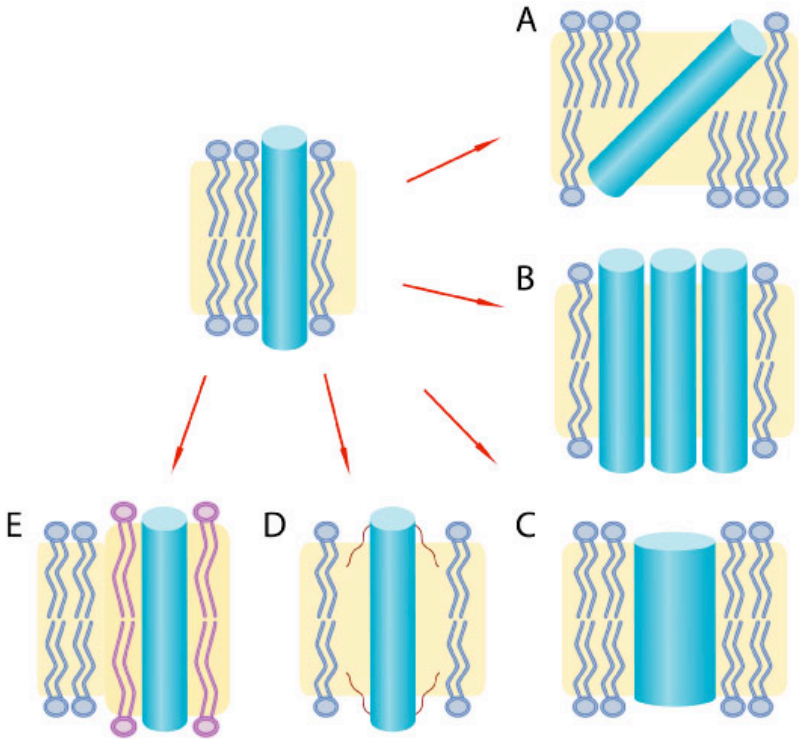


Figure 5 Different mechanisms showing how protein adapts to specific conditions in lipid bilayer (see text for details).

Many studies have indicated that transmembrane segments tilt as a response to positive mismatch, but the extent to which it occurs considerably different (Fig. 5A). Molecular dynamics studies (Kandasamy and Larson 2006) and ^{15}N NMR experiments (Harzer and Bechinger 2000) on lysine-flanked model peptides suggested

a considerable tilt as well as fluorescence studies of the natural single-span protein M13 (Spruijt, Wolfs et al. 2004). Computational studies by Venturoli and co-workers (Venturoli, Smit et al. 2005) suggested that the extent of tilt is dependent on the cross-sectional diameter of the protein and that tilting occurs more easily for single-span peptides than for large proteins. However, in this study, large proteins were represented as rigid entities, whereas in natural multispanning proteins the transmembrane segments may tilt against one another. It is possible that multispanning proteins in fact perturb lipid packing less and therefore tilt more easily than single helices. Besides tilt, oligomerization has been reported as a response to mismatch (Fig. 5B). Using different fluorescently labeled model peptides, some groups showed that the tendency to self-associate increases with either positive or negative mismatch (Sparr, Ash et al. 2005; Yano and Matsuzaki 2006), demonstrating that helix–helix association of membrane proteins can be promoted simply by less favorable helix–lipid interactions. Studies of model peptides have enabled the identification of general properties of transmembrane segments that promote helix–helix association, such as the presence of polar residues (Zhou, Merianos et al. 2001). Such studies have also greatly contributed to understanding specific dimerization motifs, such as the GxxxG motif, as first identified by MacKenzie and co-workers (MacKenzie, Prestegard et al. 1997). Proteins can also respond to changes in the lipid environment by a conformational change of their backbone (Fig. 5C). Studies on model peptides in fluid phase bilayers with different thickness suggested that systematic adaptations of the helical pitch of peptides as a response to mismatch, if any, are at most very small (de Planque, Goormaghtigh et al. 2001; Ozdirekcan, Rijkers et al. 2005). In natural membranes, cholesterol-enriched domains with different thickness occur (Simons and Vaz 2004) and proteins may be directed to such domains based on their hydrophobic length (Fig. 5E). Indeed, the length of transmembrane segments of monotopic membrane proteins is a major determinant of whether these proteins end up in the Golgi or in the cholesterol-rich plasma membrane (Munro 1995). However, matching alone is not sufficient, because some groups have reported that long model peptides do not partition into thicker liquid-ordered domains in model membrane systems (van Duyl, Rijkers et al. 2002; Vidal and McIntosh 2005).

NMR aspects of membrane proteins

Solution NMR has been established as a major method to determine structures of proteins and protein complexes in aqueous solution (Wüthrich 1986; John Cavanagh 1995). In addition to the elucidation of their structures, NMR is also used to probe dynamical processes of such proteins and complexes. Membrane proteins, when embedded in lipid bilayers, are not suitable for solution NMR techniques because their overall reorientation in these environments is slow and highly anisotropic, which leads to unfavorable relaxation properties and broadening of the resonance lines. However, solid-state NMR has been successfully employed to obtain highly resolved spectra of membrane-bound peptides and proteins in lipid bilayers. Solid-state NMR technology continues to be developed for membrane protein samples, and the structures of several small proteins in lipid bilayers have been obtained by these methods (Opella and Marassi 2004; Andronesi, Becker et al. 2005).

An alternative approach to solving structures and obtaining dynamical information on membrane proteins is to extract the proteins from their host membranes and integrate them in non-denaturing micelles, which tumble fast enough to give well-resolved resonance lines when using solution NMR methods. The membrane protein–detergent complexes are usually large compared to those protein structures that are routinely solved by NMR, so that the most advanced solution NMR techniques and spectrometers operating at highest magnetic fields are typically employed to solve structures of these membrane proteins. The accessible structural constraints for larger helical membrane protein systems are generally the same as those available for large perdeuterated globular proteins. Local backbone distance constraints can be readily obtained from amide NOEs, and phi/psi dihedral angle constraints from backbone chemical shifts (Cornilescu, Delaglio et al. 1999). Useful long-range constraints of modest precision can be derived using paramagnetic relaxation enhancement (PREs) from spin-labels specifically introduced by site-directed mutagenesis (Battiste and Wagner 2000) as described in detail for OmpA in DPC micelles (Liang, Bushweller et al. 2006). Long-range amide-amide NOEs appear to be quite rare in membrane proteins with multiple transmembrane helices. Long-range NOEs from selectively protonated methyl groups are also unlikely to be generally useful as structural constraints for these proteins – the number of methyl

groups is typically extremely large, and the chemical shift dispersion in both ^{1}H and ^{13}C dimensions is quite low, as pointed out by Sanders (Sanders and Sonnichsen 2006). However use of methyl group NOEs and chemical shift changes can be quite informative, as it was shown in the study of charybdotoxin binding to KcsA (Yu, Sun et al. 2005). RDCs measured in partially aligned samples (Tolman, Flanagan et al. 1995; Tjandra and Bax 1997; Chou, Kaufman et al. 2002; Cierpicki and Bushweller 2004; Oxenoid and Chou 2005; Cierpicki, Liang et al. 2006) are expected to play a major role in defining structures and conformational changes in large membrane proteins. Here the challenge is to induce long-term partial alignment without significantly decreasing the membrane protein concentration or increasing the linewidths of its amide resonances. Of the commonly used alignment media, only stretched polyacrylamide gels (Tycko, Blanco et al. 2000; Chou, Gaemers et al. 2001; Cierpicki and Bushweller 2004) are compatible with high concentrations of detergents or isotropic bicelles, because only in these the proteins do not integrate into the alignment media. Uncharged or charged gels have been used for partial alignment of small to medium sized membrane proteins (Chou, Kaufman et al. 2002; Oxenoid and Chou 2005; Call, Schnell et al. 2006; Cierpicki, Liang et al. 2006), but have been less successful for larger membrane protein systems (Hwang and Kay 2005; Sanders and Sonnichsen 2006; Kamen, Cahill et al. 2007) where reduced protein concentrations, interactions with the gel, and short sample lifetimes have limited their use even for measuring ^{1}H-^{15}N dipolar couplings in 2D NMR experiments. An alignment method based on DNA nanotubes was reported recently that overcomes all of these limitations (Douglas, Chou et al. 2007). The nanotubes are resistant to detergent and do not reduce achievable protein concentrations, permitting ^{1}H-^{15}N and $^{13}C_{\alpha}$-$^{1}H_{\alpha}$ dipolar couplings to be measured for two oligomeric membrane proteins. Another approach is to incorporate a lanthanide ion binding-site into the protein, and use the incorporated ion to align the protein. This can be achieved by ion binding (Veglia and Opella 2000) by addition of a metal binding motif (e.g. an EF-hand (Ma and Opella 2000), or by incorporation of single cysteines into the protein and subsequent modification with a thiol-reactive EDTA derivative (Dvoretsky, Gaponenko et al. 2002; Ikegami, Verdier et al. 2004). The thiol-reactive EDTA derivatives were found to be the most versatile method (Kamen, Cahill et al. 2007), allowing placement of the metal at different locations in the protein, typically at single cysteines that were

already engineered into the sequence for spin-label modification. Since the alignment tensor varies with different lanthanides (Barbieri, Bertini et al. 2002), multiple orientations can be obtained by varying the identity of the metal used. Pseudocontact shifts and PREs from the metals also create useful distance constraints and additionally serve to provide information on the orientation of the membrane protein within the micelle. Like DNA nanotube-induced alignment, the lanthanide method overcomes the difficulties encountered with stretched gels, permitting RDCs for multiple internuclear vectors to be readily measured for membrane proteins (Kamen, Cahill et al. 2007), with the additional advantage of being able to induce different orientations with different metals.

1.2. G-protein coupled receptors

G-protein coupled receptors (GPCRs) represent the largest family of membrane proteins in the human genome and the richest source of targets for the pharmaceutical industry. There has been substantial progress in the field of GPCR biology during the past two decades. Notable milestones include the cloning of the first GPCR genes, and the sequencing of the human genome revealing the size of the GPCR family and the number of orphan GPCRs. Moreover, there is a growing knowledge that GPCR regulation and signaling is much more complex than originally thought, and includes signaling through G protein independent pathways (Luttrell and Lefkowitz 2002; Azzi, Charest et al. 2003; Lefkowitz and Shenoy 2005). All proteins within the GPCR family contain seven highly conserved transmembrane segments (of 25–30 consecutive residues), which display a high degree of hydrophobicity. Identifying the transmembrane regions [TM1–7] also identifies the remaining structure of the GPCR (Fig. 6). This sequence will contain the three extracellular loops [EL1-3], three intracellular loops [IL1-3] as well as the protein termini. It can therefore be divided into the following regions: N terminus-TM1-IL1-TM2-EL1-TM3-IL2-TM4-EL2-TM5-IL3-TM6-EL3-TM7-C terminus (Fig. 6). The transmembrane segments form seven α-helices in a flattened two-layer structure known as the seven-helix transmembrane bundle, a structure common to all GPCRs (Yeagle and Albert 2007).

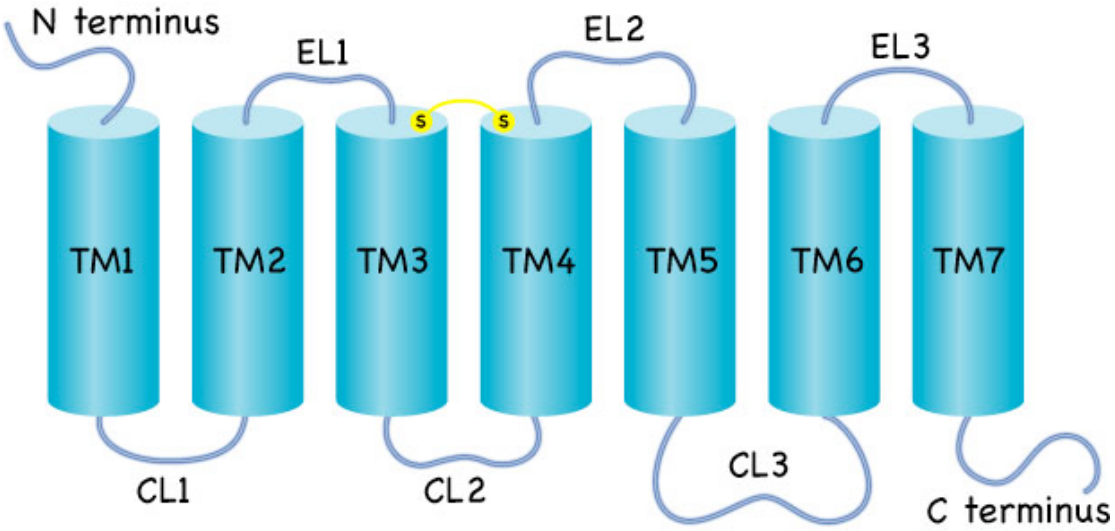


Figure 6 General structure of G-protein coupled receptors (GPCRs).

Consequently, it has been proposed that the term GPCR be used in favor of 7 transmembrane or 7TM receptors. In line with the characteristics of most related groups of proteins, the GPCR family displays a large degree of structural conservation despite little conservation of sequence. Although the remarkable advances in the biology and pharmacology of GPCRs were achieved in the last years, progress in the area of protein structure has been more limited. To date, the only high-resolution structures of a GPCR have been solved by X-ray crystallography for bovine rhodopsin in 2000 by Palczewski et al. (Palczewski, Kumasaka et al. 2000) and very recently for β2-adrenergic receptor by Rassmussen et al. (Rasmussen, Choi et al. 2007) and Cherezov et al. (Cherezov, Rosenbaum et al. 2007) (Fig. 7).

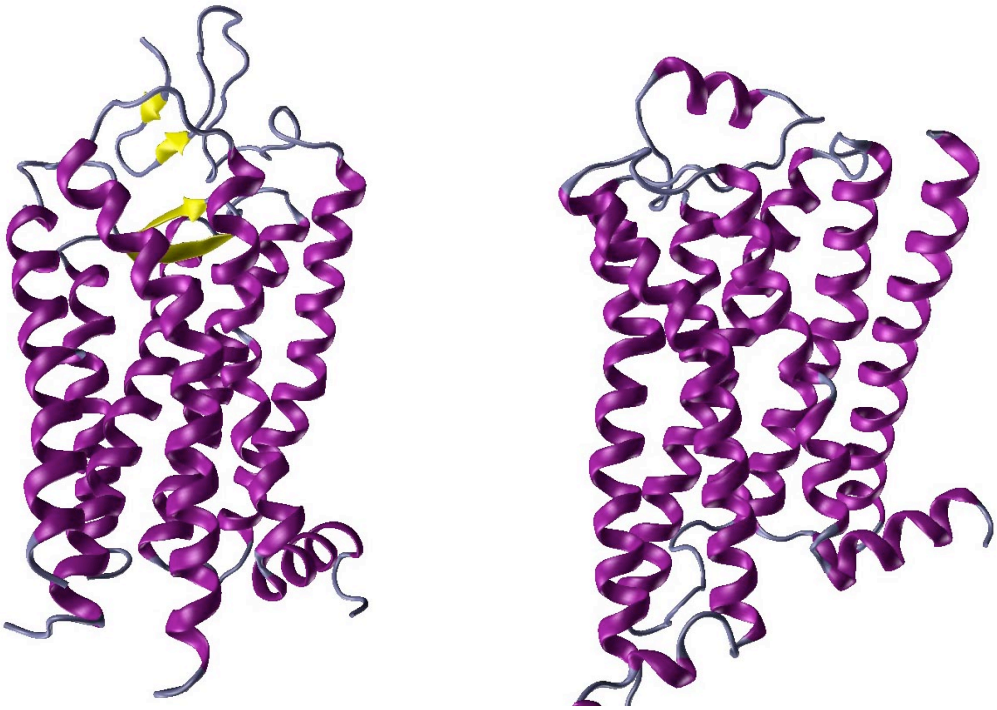


Figure 7 Rhodopsin structure (left) (Palczewski, Kumasaka et al. 2000), β2-adrenergic receptor structure (right) (Cherezov, Rosenbaum et al. 2007).

GPCRs classification

The G-protein coupled receptors (GPCRs) form a large and diverse multigene family of integral membrane proteins that are involved in many important physiological functions (Christopoulos and Kenakin 2002; Gether, Asmar et al. 2002; Bissantz 2003). GPCRs are responsible for the transduction of endogenous extracellular signals into an intracellular response. The binding of a ligand on the cell surface causes the GPCR to become activated and subsequently binds and activates ubiquitous guanine nucleotide-binding regulatory (G) proteins within the cytosol (Fig. 8(1-3)). The GPCR protein's association with the heterotrimeric G-protein complex causes the GDP bound to the Gα subunit to be exchanged for GTP (Fig. 8(4)). The Gα-GTP complex then dissociates from the Gβ subunit, releasing the Gα subunit to couple to an effector enzyme (Fig. 8(5,6)). An extremely heterogeneous set of molecules can act as GPCR ligands including ions, hormones, neurotransmitters, peptides, and proteins. Sensory GPCRs can also be activated by stimuli such as light, taste, or odor. More than one type of GPCR can interact with more than one kind of G-protein, creating a complex system involving a variety of mechanisms.

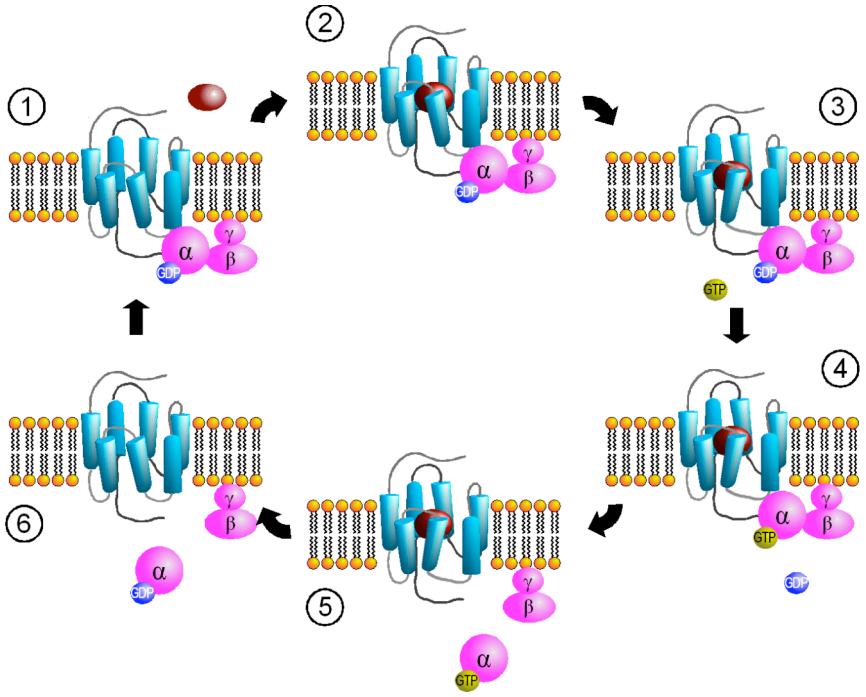


Figure 8 Scheme of the GPCR signaling cycle (see text for details).

GPCRs control and affect physiological processes such as neurotransmission, cellular metabolism, secretion, cellular differentiation, and inflammatory responses (Hebert and Bouvier 1998). Mutations in GPCR-coding genes have been linked to over 30 human diseases including retinitis pigmentosa, hypo- and hyperthyroidism, nephrogenic diabetes insipidus, as well as several fertility disorders (Schoneberg, Schulz et al. 2004). The GPCR family is a common target for therapeutic drugs and approximately 50% of all marketed drugs are targeted towards a GPCR (Klabunde and Hessler 2002). There are inherent difficulties in providing a comprehensive classification system for the GPCR family (Cheng, Carbonell et al. 2005). The term "family" has long been used to subgroup GPCRs and relies not just on sequence similarity, but also includes a corresponding set of structural, functional, and evolutionary features. Evolutionary relationships between different GPCR groups are not certain; some receptors may have arisen through convergent evolution to adopt a particular structural scaffold, and may not be homologous. One of the first GPCR family classification systems was introduced by Kolakowski (Kolakowski 1994). GPCRs were divided into seven families, designated A–F and O, derived from original standard similarity searches. This system was further developed by Horn et al. (Horn, Bettler et al. 2003) for the GPCR database (GPCRDB) database. The GPCRDB database divides the family into six classes. These are the Class A

rhodopsin-like, which account for over 80% of all GPCRs, Class B secretin-like, Class C metabotropic glutamates, Class D pheromones, Class E cAMP receptors and the Class F frizzled/smoothened family.

Class A is the largest of the human GPCR subtypes. There are at least 286 human non-olfactory Class A receptors, the majority of which is activated by photons and a variety of ligands such as odorants and other small molecules, peptides and large glycoprotein hormones (Fridmanis, Fredriksson et al. 2007) (Fig. 9(1)). A common structural feature of the GPCRs of this family is a disulfide bridge between the extracellular loop EL2 and the C-terminal end of the EL1 loop. The receptors binding endogenous peptides have an import role in mediating the effects of a wide variety of neurotransmitters, hormones, and paracrine signals. The receptors that bind biogenic amines, e.g., norepinephrine, dopamine and serotonin, are very commonly modulated by drugs. Pathological conditions, including Parkinson's disease, schizophrenia, drug addiction, and mood disorders are examples of where imbalances in the levels of biogenic amines cause altered brain functions.

Class B receptors bind the large peptides such as secretin, parathyroid hormone, glucagon, glucagon-like peptide, calcitonin, vasoactive intestinal peptide, growth hormone releasing hormone, and adenylyl cyclase activating protein (Cardoso, Pinto et al. 2006) (Fig. 9(2)). The members of this class of GPCRs share morphological features similar to the GPCRs of Class A despite their low sequence homology, however they lack the characteristic disulfide bond. Metabotropic glutamate receptors (mGluRs), forming Class C, are activated through an indirect metabotropic process. Like all glutamate receptors, mGluRs bind to glutamate, an aminoacid functioning as an excitatory neurotransmitter. In humans, mGluRs are found in pre- and postsynaptic neurons in synapses of the hippocampus, cerebellum, and the cerebral cortex, as well as other parts of the brain and in peripheral tissues. In contrast to the GPCRs of the first two classes whose ligand binding pockets are formed by their transmembrane helices and/or extracellular loops, the binding to Class C receptors occurs in a large extracellular domain, the so-called ligand-binding region (Fig. 9(3)). Pheromones targeting class D receptors are used by organisms for chemical communication (Das and Banker 2006) and cAMP receptors (class E) are part of chemotactic signaling systems (Nakagawa, Sakurai et al. 2005). Frizzled receptors (class F) are necessary for Wnt binding while the smoothened receptor

mediates hedgehog signaling (Gloriam, Schioth et al. 2005; Prabhu and Eichinger 2006). The six different classes can further be divided into subfamilies and sub–subfamilies based upon the function of the GPCR protein and the specific ligand that it binds.

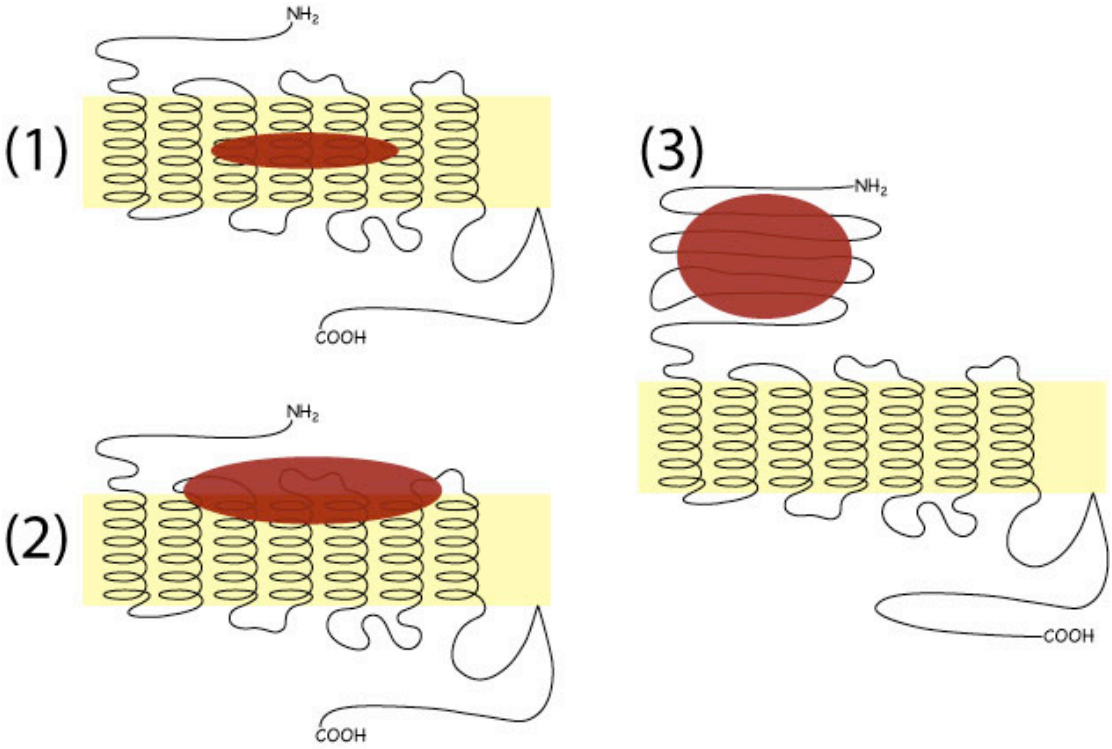


Figure 9 GPCR classification based on ligand binding site: (1) Monoamine ligands, rhodopsin; (2) Peptide ligands; (3) Glycoproteins, glutamate, Ca^{2+}, GABA ligands.

Activation of GPCRs

The understanding of signal transduction via G-protein-coupled receptors has developed dramatically during the last two decades. The initial idea of linear signaling pathways transferring information from the cell membrane to the nucleus has evolved into a complicated network of signaling pathways offering the possibility of crosstalk, fine tuning and specific regulation at multiple levels. G protein-coupled receptors (GPCRs) were initially regarded to adopt an inactive and an active conformation and to activate a single type of G protein (Wess 1998). Studies with recombinant cell systems have led to a more complex picture. First, GPCRs can activate distinct G protein species. Second, GPCR multistate models have been invoked to explain their complex behavior in the presence of agonists, antagonists and other binding partners.

The molecular mechanism of ligand activation is understood best for rhodopsin and related visual pigments. These contain the covalently bound light-sensing chromophore, 11-cis-retinal, which is an inverse agonist (Meng and Bourne 2001; Teller, Okada et al. 2001). In general, binding of an inverse agonist causes the transition of a GPCR from the native, partially active state to an inactive state (Gether 2000). Indeed, the opsin form of rhodopsin is partially active and the rhodopsin form is inactive (Meng and Bourne 2001). Rhodopsin, once excited by light, cycles through a series of spectral intermediates that produce metarhodopsin I. During this process nearly all the changes in the receptor are limited to the ligand with few changes in the protein structure. It is only upon moving to metarhodopsin II, the next step after metarhodopsin I that the receptor becomes active, the conformation changing such that the receptor can bind the G protein and activate it. The ligand, 11-cis-retinal, changes position in the protein upon activation to metarhodopsin II. The retinal translates in the helical bundle in the direction of TM5 with a simultaneous rotation of the C20 group of retinal (Patel, Crocker et al. 2004). Consequently the contact between C20 of the retinal and W265, found in the ground state, is lost and a contact of W265 with the C19 methyl group is substituted (Crocker, Eilers et al. 2006). The ring of the retinal was reported to contact TM helix 4 of rhodopsin upon activation, a contact that is not possible in the ground state structure (Borhan, Souto et al. 2000). Site-directed spin labeling experiments provided important site-to-site distances within the helical bundle of the GPCR. From such experiments, investigators have discovered that the helices in the transmembrane domain change their orientation relative to each other when metarhodopsin II is formed. For example, TM2 of rhodopsin moves relative to helix 8 upon activation to metarhodopsin II (Altenbach, Klein-Seetharaman et al. 2001), as does TM6 (Altenbach, Cai et al. 1999). The movement of TM6 was described as a rigid body movement (Farrens, Altenbach et al. 1996). However, studies of the influence of prolines on helix conformation indicated that the helical segments on either side of the proline can turn with respect to each other and this can be a source of some of the conformational changes seen in the transmembrane domain (Cordes, Bright et al. 2002). TM7 and TM1 become further separated (Altenbach, Cai et al. 2001) upon activation. Data such as those derived from site-directed spin labeling, provided some of the few details known about the structural changes that occur in the protein upon activation to metarhodopsin II. The

same approach was used to determine a structure for ground state rhodopsin from sparse distance constraints resulting the structure for metarhodopsin II (Choi, Landin et al. 2002). This structural determination exploited the changes in distances from site-directed spin labeling to observe the structural changes that occurred in three dimensions when metarhodopsin II. The greatest density of long-range distance constraints was in the cytoplasmic face of the receptor, the surface that couples to the G protein. The most obvious change in the cytoplasmic face of this receptor is the opening of a cleft in the surface. This was found to result from a breaking of the ionic lock between R135 of the highly conserved DRY sequence at the end of helix 3, and E247 on helix 6. This ionic interaction stabilizes the ground state and apparently must be broken to form the excited state. The cleft that appears interacts with portions of the G protein, as had been suggested previously (Meng and Bourne 2001). Using studies on fragments of the G protein binding to metarhodopsin II and the structure derived for metarhodopsin II (Choi, Landin et al. 2002), as well as other experimental clues to the interface, a model for binding of the G protein, transducin, to metarhodopsin II was described (Yeagle and Albert 2003). The hypothesis was advanced that the G protein bound through an induced fit mechanism in which the binding energy of transducin to metarhodopsin II was utilized to induce a conformational change in the G protein.

In the amine receptors, agonist binding disrupts the salt bridge between TM3 and TM6, similar to opsins (Porter, Hwa et al. 1996). In the peptide-hormone receptors salt-bridge disruption is not common, but displacement of residues in TM3 leads to activation (Ji, Grossmann et al. 1998; Miura, Feng et al. 1999). Disulfide cross-linking between cysteine pairs introduced artificially into TM3 and TM6 prevents G-protein activation by the cytoplasmic loops. The magnitude of movement of TM6 is greater than that of TM3 and TM7. As a result, the inner faces of TM2, TM3, TM6 and TM7 become more exposed and the cytoplasmic ends of TM4 and TM5 become less exposed (Meng and Bourne 2001). The initial ligand recognition appears to be quite different for class C GPCRs (Pin, Galvez et al. 2003). Here, the natural agonist binding sites are contained within the large extracellular N-terminal domain of the receptor. After binding of the messenger, this N-terminal domain is thought to “close“ over the receptor surface. The second part of the activation process is likely to involve the TM domains of class C GPCRs as well. This allows small

antagonistic molecules to intercalate within the receptor central cleft and to block the activation process in a non-competitive fashion (Carroll, Stolle et al. 2001).

GPCRs oligomerization

The hypothesis of GPCR dimerization or oligomerization was first suggested in the 1970s from cooperative ligand binding and radiation inactivation studies (Park, Filipek et al. 2004). However, the concept of GPCR oligomerization gained little attention because such early evidence was equally consistent with other hypotheses. Furthermore, as GPCRs were characterized molecularly, there was no obvious mechanistic reason why they would need to function as oligomers in order to activate G proteins. Indeed, biochemical and biophysical studies of rhodopsin in native membranes or solubilized in detergent indicated that the monomeric form of this GPCR can activate its cognate G protein, transducin (Chabre and le Maire 2005). Nevertheless, extensive evidence accumulated over the past decade indicates that many GPCRs oligomerize in living cells, that various types of GPCRs can hetero-oligomerize, and that oligomer formation is critical for receptor biogenesis and function (Bouvier 2001; George, O'Dowd et al. 2002; Park, Filipek et al. 2004; Parnot and Kobilka 2004; Park, Lodowski et al. 2008).

Organization of GPCRs into oligomeric clusters has been discovered from studies using several different methods, including radioligand binding, coimmunoprecipitation, bioluminescence resonance energy transfer (BRET), fluorescence resonance energy transfer (FRET), and atomic force microscopy (AFM). Cryo-electron tomography images of minimally perturbed murine rod outer segments display a highly concentrated heterogeneous distribution of rhodopsin in disc membranes (Nickell, Park et al. 2007). AFM images of rhodopsin in native disc membranes provide the clear structural picture of the oligomeric arrangement of a GPCR (Fotiadis, Liang et al. 2003). Rows of rhodopsin dimers have been observed directly by AFM and the constraints from these studies have been used to construct a molecular model of a rhodopsin oligomer (Fotiadis, Liang et al. 2004). The model indicates that the rhodopsin dimer offers a complementary platform for the binding of a single transducin molecule (Filipek, Krzysko et al. 2004).

The dimer interface of GPCR oligomers likely involves the transmembrane regions of the receptor. The large amino-terminal regions that form ligand-binding domains in some GPCRs may also form a dimer interface because those regions have a propensity to form oligomers, even in the absence of the transmembrane region of the receptor. The nature of the dimer interface specifies not only which GPCRs can exhibit productive interactions, but influences potential allosteric interactions between dimer partners. Although many GPCRs have been shown to participate in homodimerization, heterodimerization is more variable, with some GPCRs exhibiting broad promiscuity (Salim, Fenton et al. 2002) and others exhibiting a high degree of selectivity (Kroeger, Hanyaloglu et al. 2001). Experimental data suggest that there are diverse dimerization interfaces or domains both within the transmembrane helices (Overton and Blumer 2002) as well as at either the extracellular amino terminus (Pace, Gama et al. 1999) or the intracellular carboxyl terminus (Calver, Robbins et al. 2001) depending on the GPCR.

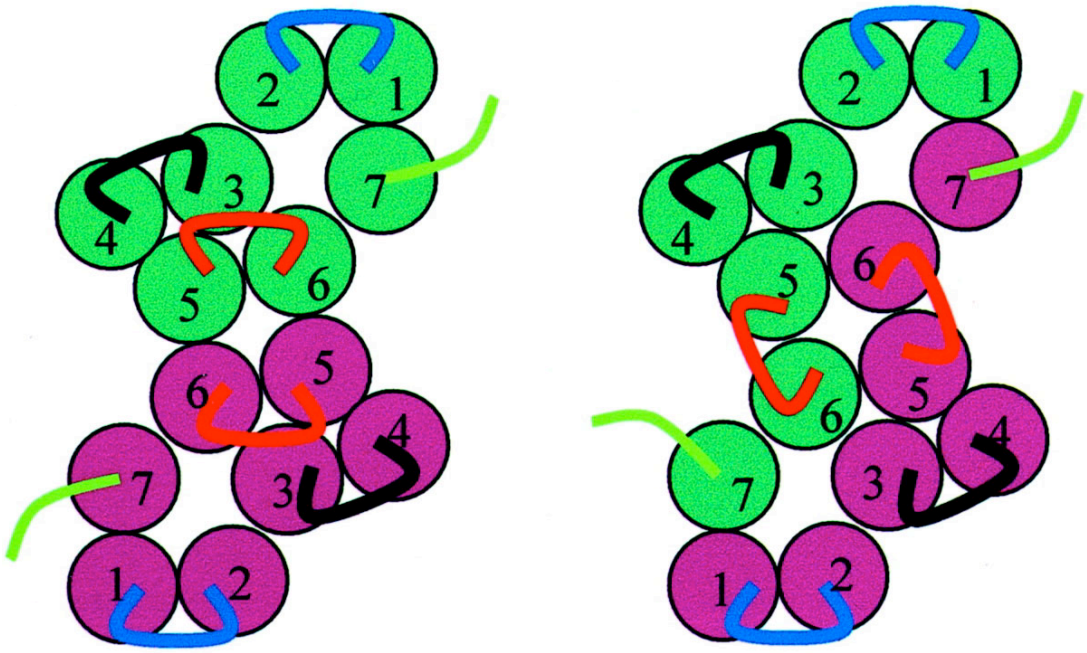


Figure 10 Potential GPCR dimer interfaces. Contact dimers (left): the interface between GPCR monomers involves surface contact between helices of two independent monomers (indicated by color). Domain-swapped dimers (right): helices 6 and 7 are exchanged or "swapped" between GPCR monomers (indicated by color) (Breitwieser 2004).

Two modes of interaction have been described (Fig. 10): (1) contact dimerization, in which the relevant helices from one monomer contact partners in the other monomer, stabilizing the dimer pair; (2) domain swapping, in which several helices from each receptor are "swapped" in the dimer, such that the functional monomer within the dimer contains helices contributed by both receptors (Fig. 10).

Contact sites have been identified for a very few GPCRs and mostly for homodimers. Synthetic peptides related to TM6 block both dimerization and activation of the $ß_2$AR (Hebert, Moffett et al. 1996) whereas TM4 has been shown to mediate homodimerization of D2 receptors by cysteine scanning mutagenesis and chemical crosslinking (Guo, Shi et al. 2003) and C5a receptors by disulfide trapping (Klco, Lassere et al. 2003). TMs 1 and 2 were shown to be involved in yeast α-factor receptor dimerization by FRET analysis (Overton and Blumer 2002). A computational subtractive correlation method (based on the rhodopsin crystal structure, solvent accessibility, and location of residues on outer faces of helix bundles) has been applied to opioid receptor homodimers and heterodimers, and results indicated a high degree of variability in interaction domains (Filizola and Weinstein 2002). The domain-swapping model was first suggested by the functional rescue upon coexpression of m3/α_{2c} and α_{2c} /m3 chimeric receptors, because binding of both m3 and α_{2c} agonists was restored (Maggio, Vogel et al. 1993). Domain swapping has also been proposed as an explanation for the different ligand binding specificity of coexpressed δ and κ opioid receptors (Jordan and Devi 1999) because different affinities can result when a functional binding unit comprises helices from two distinct receptor types (Fig. 10). Although there is clear evidence for domain swapping in specific cases (Dean, Higgs et al. 2001), systematic study of several other GPCRs has revealed no evidence for this mechanism. The domain swapping may represent only one of several mechanisms utilized by particular GPCRs in dimer/oligomer formation and stabilization. It has been suggested that contact dimers and domain-swapped dimers have equivalent abilities to signal to G proteins, and the ability of a particular GPCR pair to form a functional dimer via either interaction depends on the relative energetics of the two possible pairings (Dean, Higgs et al. 2001).

GPCR fragments

Despite recent promising attempts to incorporate intact GPCRs into micellar systems (Tian, Breyer et al. 2005; Park, Prytulla et al. 2006; Tian, Breyer et al. 2006), direct analysis of these molecules by solution NMR is at present possible to only a limited extent due to large problems associated with producing full-length, biologically active receptors, purifying and reconstituting these molecules, and measuring high quality NMR spectra in membranes. Taking into account the enormous problems associated with the expression, purification, and reconstitution of intact GPCRs into membrane-mimetic environments suitable for biophysical studies, investigations on fragments of these proteins seem to be well justified. As the size of these fragments increases to include more than one helix, the understanding of the involvement of helix-helix interactions in influencing the secondary and tertiary structure of individual TM helices will improve. Although many synthetic problems may be reduced when using fragments of GPCRs, a crucial issue remains whether these truncated constructs are able to successfully mimic structural features of the much longer polypeptides.

One question that arises is the extent to which structures of receptor fragments reflect the secondary structure of the native protein. This question has been directly addressed using bacteriorhodopsin. A series of peptide fragments of the protein were designed, each overlapping their neighbor in the sequence by about 10 residues. Each fragment represented either a turn or a transmembrane helix of bacteriorhodopsin. High-resolution NMR structures were obtained from each fragment and the structures were overlaid on the corresponding part of the X-ray crystal structure. Good agreement was observed between helices in the protein and the structures of the fragments, and turns in the protein and the structures of the fragments (Katragadda, Alderfer et al. 2000; Katragadda, Alderfer et al. 2001). Extensive studies have led to the conclusion that when a fragment of a membrane protein exhibits secondary structure, that structure is similar to the structure in the intact protein. In the case where the fragment is disordered, no conclusion can be made.

Another question comes from the study of loops of GPCRs as fragments: will peptide fragments containing the sequences of loops form loop structures without constraints on the position of the amino and carboxyl termini? In some cases, loops naturally formed with free amino and carboxyl termini. In these cases, the residues in

the turn likely directed the form of the structure, perhaps enhanced by helix–helix interactions when the structure was defined by a helix–turn–helix motif (Ulfers, Piserchio et al. 2002; Demene, Granier et al. 2003). In some examples, β-turns formed which were stabilized by internal hydrogen bonding (Yeagle, Alderfer et al. 1997). In other cases, covalently linking the two ends of the turn enhanced the stability of the turn (Ruan, So et al. 2001). Finally, structural studies of loops in the presence of detergent micelles as a membrane mimetic also stabilized the structures (Ulfers, Piserchio et al. 2002). Less unusual the peptide fragments containing sequences of transmembrane helices of GPCRs formed stable helices separate of the remainder of the protein (Lazarova, Brewin et al. 2004).

Many groups have demonstrated that smaller fragments of soluble, well-structured proteins display increased propensity to transiently adopt conformations similar to those encountered for that particular stretch when placed in the context of the full polypeptide (Dyson, Rance et al. 1988). However, these studies have usually also revealed that the fragments are still fairly flexible mostly because solvation competes with intramolecular hydrogen bonding. Solvation is much less favorable when the polypeptide is partitioned into a membrane, and water is essentially excluded from the membrane interior (Popot and Engelman 2000). Indeed, many peptides, which are unfolded in water, adopt secondary structure when placed into a membrane-mimicking environment, known as the coupled partitioning folding (White and Wimley 1999). Furthermore, many relatively short peptides fold into stable helices in micelles, both in transmembrane (Papavoine, Aelen et al. 1995; MacKenzie, Prestegard et al. 1997) or surface-associated fashion (Brown and Wuthrich 1981; Ladokhin and White 1999; Bader, Bettio et al. 2001). Yeagle and co-workers have synthesized peptides corresponding to the cytosolic loops and TMDs of rhodopsin and bacteriorhodopsin and studied them by solution NMR, proposing that the structures of these fragments resemble the corresponding regions in the native receptors (Yeagle, Alderfer et al. 1995; Yeagle, Alderfer et al. 1995; Yeagle, Alderfer et al. 1997; Yeagle, Alderfer et al. 1997). Pellegrini and Mierke studied the extracellular domain of the PTH1 receptor in the presence of DPC micelles (Pellegrini and Mierke 1999). Recently much progress has been made in expressing and isotopically labeling regions of fragments of Ste2p (Estephan, Englander et al. 2005) and CB2 (Zhao, Zheng et al. 2006; Zheng, Zhao et al. 2006) receptors containing loops and up to two TMs of these GPCRs. The structure and topology in the micelle of 73-residue peptide

containing TM7 from Ste2p was investigated by solution NMR methods (Neumoin, Arshava et al. 2007) resulting a kinked around internal Pro residue α-helix for the transmembrane segment and a low-resolved nascent helix in the N-terminal part. For the CB2 receptor 54-residue and 74-residue peptides no high-resolution structure yet available.

The work described above on peptide fragments of GPCRs derived from transmembrane segments of the protein revealed that the transmembrane helices of GPCRs are intrinsically stable in their helical secondary structure, a stability apparently arising from their amino acid sequences. These results raise the question of the role of helix–helix interactions in GPCRs. Helix–helix interactions have been discovered that contribute to the stability of GPCRs. Co-expression of bundles of transmembrane helices of rhodopsin (3 TM and 4 TM, or 5 TM and 2 TM) can result in correct reassembly of the fragments into a functional pigment (Yu, Kono et al. 1995). This remarkable result identifies a strong and specific interaction within the bundle of transmembrane helices. Such helix–helix interactions were studied in detail for the adenosine A2A receptor. These studies reported that helix stability was enhanced in some cases by the presence of neighboring helices during protein folding (Thevenin, Roberts et al. 2005). These authors also found evidence for self-association of transmembrane helix 5, which could be part of the foundation for dimerization of receptors (Thevenin, Lazarova et al. 2005).

These studies suggest that stability and folding of GPCRs are influenced by several factors. The helical transmembrane segments of GPCRs have considerable local stability and will spontaneously fold, likely early in the folding process (White and Wimley 1999). Helix–helix interactions are specific and strong as reflected in the ability of separately expressed bundles of GPCR transmembrane helices to associate correctly. Loops connecting the transmembrane helices in many cases exhibit intrinsic stability and that also contributes to overall GPCR stability. Therefore the stability of GPCRs is built from short-range intrinsic stability of secondary structure (helices and turns) and from helix–helix interactions within the membrane bilayer.

1.3. Folding of GPCR ligands in solution

Most of the proteins fold into a unique three-dimensional (3D) structure to be biologically active, however proteins are synthesized as linear chains of amino acids on ribosomes. Each newly synthesized chain has to search its way to a unique, active conformation following its synthesis. In simple terms, folding could be described as the process by which the many degrees of freedom existing in unfolded polypeptide chains become coordinated into well-defined structures through energetics specific to the corresponding amino acid sequences. An understanding of the protein folding process is of enormous importance, because it will provide the "link" in the flow of information between a gene sequence and the 3D structure of a protein.

Pathways of GPCR-peptides interaction

Among several classes of GPCRs, receptors that use endogenous peptides as ligands are known as peptidergic GPCRs (Rashid, O'Dowd et al. 2004). This group is diverse and at least 35 different families and their ligands have been identified (Sankararamakrishnan 2006). Majority of peptidergic GPCRs fall into the category of class A (or rhodopsin-like) GPCRs and the rest can be classified as class B (secretin family). In peptide hormones, residues that govern specificity and possess activity form "message" segment. The residues of the message segment are evolutionarily conserved for a given peptide family and are responsible for triggering all the receptors of that family. The variable region directs the message to the individual receptor subtypes within a family and is called "address" segment. Message segment may lie either in the C-terminal or N-terminal end of peptide hormones. Conformational features of endogenous peptide ligands could control receptor binding/selectivity and influence their biological activity. Short and linear endogenous peptide hormones are usually very flexible and have been shown to assume random conformations in aqueous medium (*vide infra*).

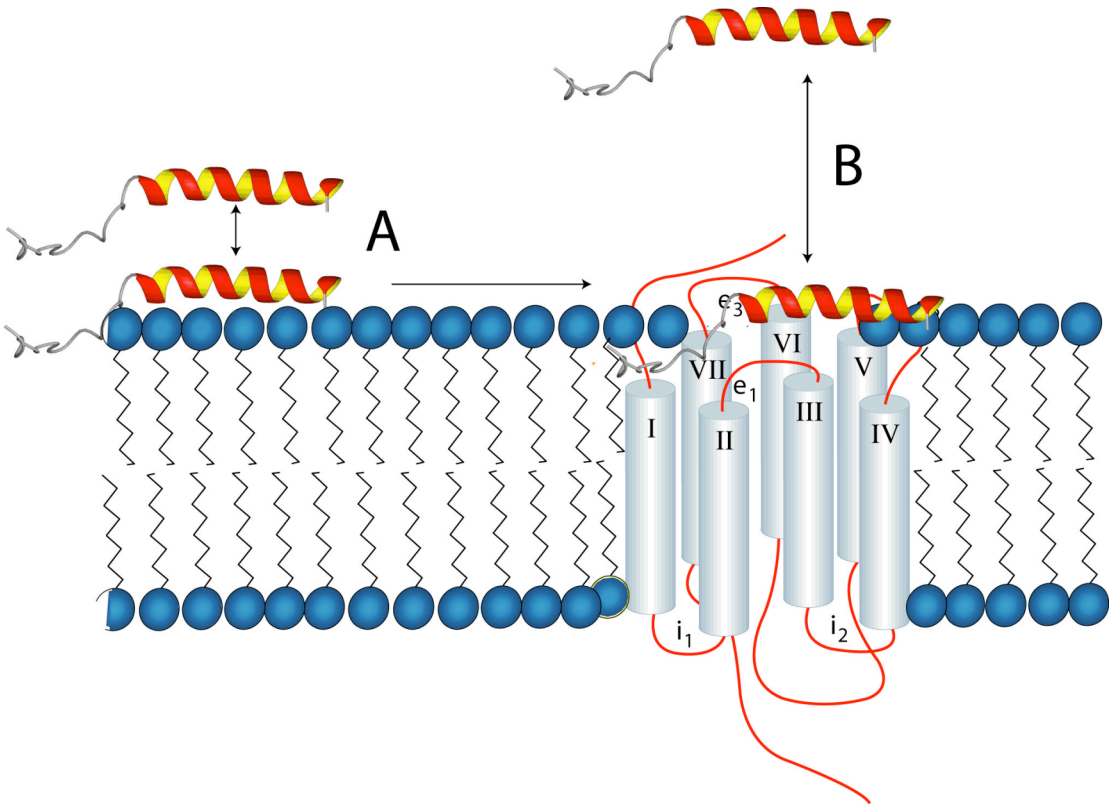


Figure 11 Schematic representation of ligand binding to GPCR. A – membrane-assisted binding; B – direct binding.

Peptide hormones can interact with its receptor site in at least two ways (Fig. 11). There could be a direct interaction from the aqueous phase with the extracellular loops of GPCRs. There is also a possibility of pre-adsorption of peptide hormones to the target cell membranes followed by subsequent interaction with the receptor. Kaiser and Kezdy studied biologically active peptides that act on cell surfaces and demonstrated the role of amphiphilic cell surface environment in influencing the structure of the peptides (Kaiser and Kezdy 1983). By increasing the amphiphilicity of the structurally important regions of the molecules that is complementary to the cell surfaces, biological activity of the peptides could be enhanced. Models of peptide hormone neuropeptide Y were designed by Minakata et al. (Minakata, Taylor et al. 1989) to investigate the role of the hydrophobic and hydrophilic domains of potential amphiphilic α-helix in this peptide. Their experiments demonstrated that amino acids in the hydrophobic side of the amphiphilic α-helix could be replaced without much loss in activity provided that the hydrophobicity of amino acids was not changed. McLean et al. synthesized cyclic, conformationally restricted disulfide analogs of NPY to investigate the role of the amphipathic helix (McLean, Buck et al. 1990). The

synthesized peptides contained various lengths of amphiphilic helical region of the peptide. It was shown that the peptide with larger amphiphilic helical region had significant interactions with lipids and higher potency in pig spleen receptors. Robert Schwyzer and his colleagues studied several regulatory peptides on the surface of artificial lipid bilayer membranes and developed the concept proposed by Kaiser and Kezdy (Kaiser and Kezdy 1983). The peptide compounds were found to interact with the bilayers (Schwyzer 1991) and the resulting conformational preferences in the membrane-bound state were correlated with their bioactivities. While conformational space of the short peptide is generally poorly defined in aqueous solution, it is strongly affected by the local environment. Based on these studies, Schwyzer postulated "membrane compartments theory", a two-step model for the peptide–receptor interaction (Schwyzer 1991; Schwyzer 1995). In this concept, it is the membrane-bound conformation of the ligand that is recognized by the receptor (Fig. 11A). In the first step, it is proposed that cell membrane induces preferred conformations and orientations of the peptide by guiding important residues into different compartments of the cell membrane (hydrophobic, interface or bulk water environments). In the second step, the peptide undergoes two-dimensional lateral diffusion on the membrane surface to the receptor where receptor recognition and binding occur. It should be pointed out that the concept of the membrane-bound pathway of receptor recognition as developed by Schwyzer does not exclude the possibility of conformational changes both in ligand and the receptor resulting from intermolecular contacts that are established during the recognition event. Experimental support for Schwyzer's theory came from the work of Moroder et al. on cholecystokinin peptide (Moroder, Romano et al. 1993). A fully active CCK analogue was covalently linked to 1,2-dimyristoyl-3-mercaptoglycerol. It was demonstrated that the lipophilic CCK adduct inserted rapidly into phospholipid bilayers. Binding experiments showed that lipo-CCK derivative competes with the unmodified CCK compound for the receptor binding and this confirmed a two-dimensional membrane-bound migration of the ligand to the receptor.

Protein folding models in solution

The modern views of folding had been set over 30 years ago (Anfinsen 1973). Progress was remarkable even with the limitations of experimental and theoretical

methods at that time. It was clear that protein folding is a reversible and spontaneous event. In addition, kinetic experiments had already been performed showing that folding occurs rapidly, usually within a few seconds. In contrast, simple calculations suggested that even a small protein of 100 amino acids would take over a billion years to find its native state if folding were to occur by a random search of all possible conformations (Levinthal 1968). These observations led to the view that there must be a specific pathway for folding, and a crucial issue was to understand how a protein finds the "right" pathway and avoids the others. In addition, it was revealed that metastable, partially folded states of proteins exist and, if they could be detected and studied, these species might provide important insights about the transition to the native state. Anfinsen's original experiments demonstrated that proteins fold spontaneously and reversibly into their native conformation (Anfinsen 1973). The nucleation growth model proposed that residues adjacent in sequence form a nucleus from which the native structure then develops in a sequential manner (Fig. 12). By contrast, the framework model suggested that local elements of secondary structure form first and that these then dock into the native tertiary structure of the protein, possibly by a diffusion–collision mechanism (Fig. 12). In the hydrophobic collapse model, a protein buries its hydrophobic side chains from solvent water early during folding, forming a collapsed intermediate or molten globule species, from which the native state develops by searching within this conformationally restricted state. Finally, the jigsaw model suggests that each protein molecule could fold by a different path (Kim and Baldwin 1982). Major advances have been made in elucidating the folding mechanisms of proteins since the original models described above. These have been derived from new and powerful experimental approaches, combined with the development of solid theoretical methods.

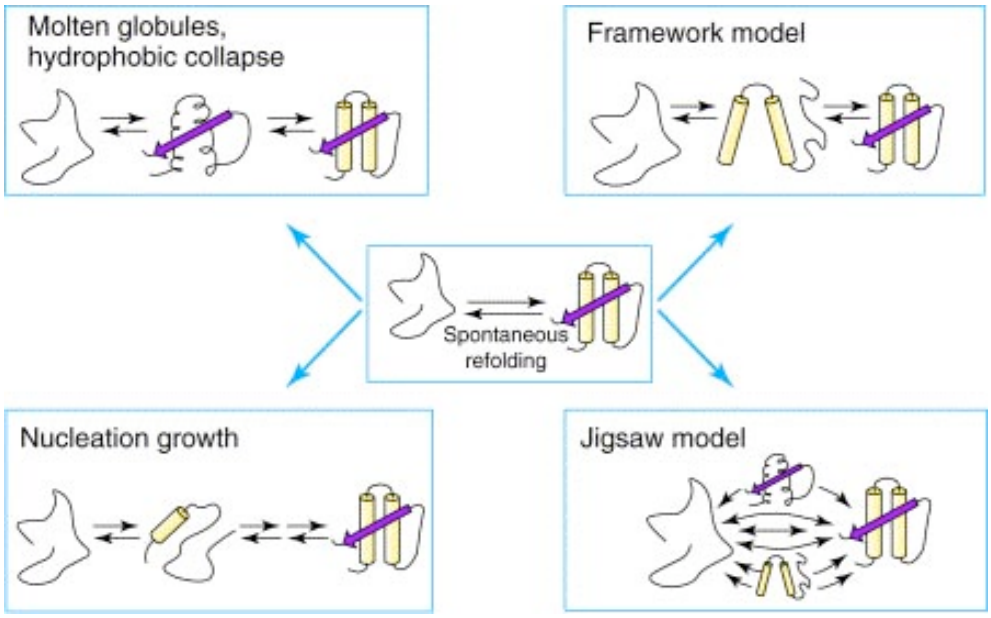


Figure 12 Different pathways of protein folding (see text for details).

One of the most important ideas in the theory of folding today is that there is not a single, specific folding pathway, as was suggested in some early models (Kim and Baldwin 1982). Instead, a multidimensional energy landscape or folding funnel better describes the folding process (Fig. 13). Thus, there are potentially many routes to the native state and which pathways are populated will depend on the details of the system being studied (e.g. the amino acid sequence, the topology and the experimental conditions). Different routes might be populated and/or different intermediates and transition states observed as a consequence of relatively small alterations of a common free-energy profile. A good starting point when studying folding is to choose simple, small proteins as model systems.

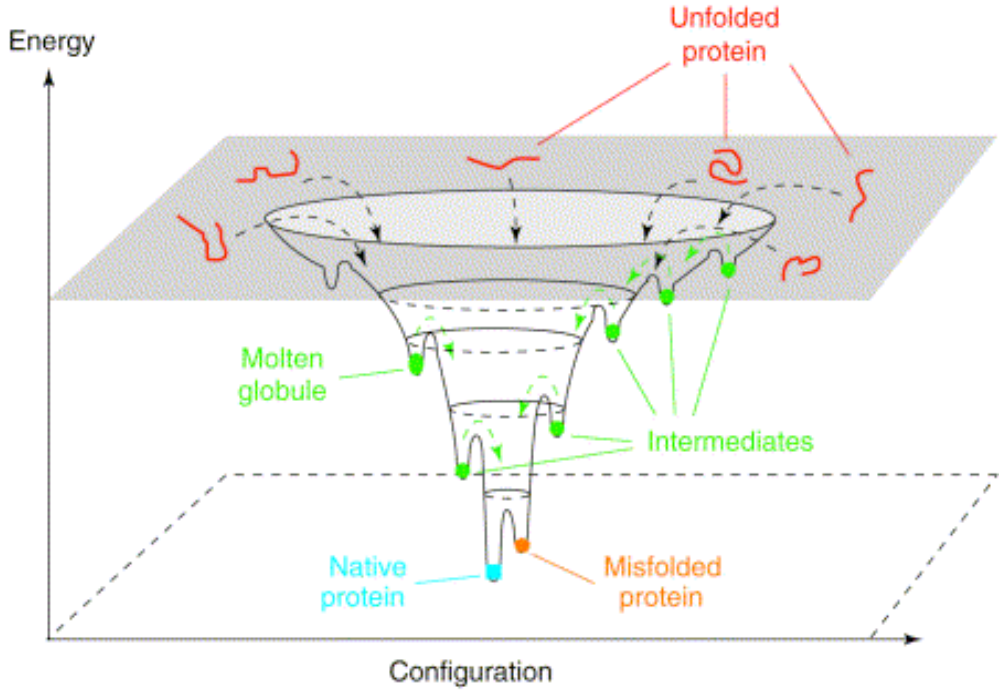


Figure 13 Cartoon illustrating the protein folding funnel.

The folding of many of small proteins or isolated domains with less than 100 amino acids have now been studied and these have provided some of the new insights into folding mechanisms (Jackson 1998). The first of these proteins to be studied was chymotrypsin inhibitor 2 (Jackson and Fersht 1991). This simple 64-residue protein was found to fold without populating intermediates in a two-state transition manner. Similar behavior has been observed for many small, single-domain proteins (Jackson 1998). Two-state systems are the simplest models of protein folding. In these cases, only the unfolded state and the folded, native state are populated on the folding pathway. Both unfolding and refolding in this case are monophasic processes.

Recent theoretical and experimental results reveal the existence of small barriers to protein folding. These barriers can even be absent altogether, resulting in a continuous folding transition, in other words downhill folding (Munoz and Sanchez-Ruiz 2004). One of the first downhill folding proteins studied was the 40-residue protein BBL from Escherichia coli that unfolds atom by atom starting from a defined three-dimensional structure as was shown by monitoring NMR chemical shifts during thermal unfolding by Sadqi et al. (Sadqi, Fushman et al. 2006).

NMR methods used to study protein folding

NMR spectroscopy is a particularly well-suited technique to address the important issue of protein folding, because it can be used to probe structure in highly dynamic systems and because site specific probes of the formation or removal of structure are available through, for example, hydrogen exchange experiments (Bai, Sosnick et al. 1995). Molten globule states of proteins often representing folding intermediates, are characterized by some extent of secondary structure, particularly α-helices. For example, the molten globule of α-lactalbumin at pH=2 has a near-native helical content (Arai and Kuwajima 2000). Hydrogen–deuterium exchange methods monitored by 2D NMR can be used to identify the specific residues involved in the regions of secondary structure. If a protein is dissolved in D_2O, the backbone amide protons will exchange with deuterons. The intrinsic kinetics of the exchange process depends on pH, temperature, and the neighboring side chains (Krishna, Hoang et al. 2004). If an amide is involved in a hydrogen bond then the rate of exchange may be slowed down significantly compared to the predicted intrinsic rate. The level of protection of an amide from hydrogen–deuterium exchange, the protection factor (PF), is measured as the ratio of the intrinsic exchange rate (k_{intr}) to the observed exchange rate (k_{obs}), $PF=k_{intr}/k_{obs}$. In native proteins, protection factors of 10^6 are common in regions of hydrogen-bonded secondary structure. In completely unfolded proteins, the observed rates of exchange are those predicted from the intrinsic rates. Molten globules contain significant levels of secondary structure and the hydrogen bonds present in these can give rise to protection from hydrogen–deuterium exchange. The poor resolution and extreme line broadening often observed in the NMR spectrum of a molten globule at low temperature mean that the slowly exchanging amides located in regions of secondary structure cannot be identified directly from the NMR spectrum of the molten globule. Instead, the increased resolution and narrow lines observed in the native state spectrum of the protein are exploited.

A major source of information about the protein folding intermediate states can be obtained using ^{15}N relaxation measurements, which are sensitive to motions of the backbone amides. ^{15}N R1 and R2 relaxation rates and the $^{15}N\{^1H\}$ steady-state heteronuclear NOE are most common parameters used for characterizing the folding process. The $^{15}N\{^1H\}$ NOE is sensitive to motions on the picosecond to nanosecond

time scale. Negative NOEs are characteristic of the unstructured regions of the protein that behave like a free-flight chain, whereas in structured regions of the protein the observed NOE ratio is around 0.6 (Eliezer, Chung et al. 2000). R2 is sensitive to nanosecond motions and to exchange processes on the millisecond to microsecond time scale. Higher R2 rates are observed in the structured regions of the protein indicating the dominance of slower time scale motions. These R2 values may include contributions from slower exchange processes resulting from motions of the structural elements within the protein core (Eliezer, Chung et al. 2000).

Spin labeling is a useful approach for the identification of long-range contacts in unfolded and partially folded proteins. The presence of a label containing a free radical, covalently attached to a unique cysteine residue introduced by mutagenesis will lead to a perturbation of the relaxation rates of nuclei that are located within up to 20 Å radius of the spin label (Gillespie and Shortle 1997). This interaction can be detected by comparison of R1 and R2 rates measured in the presence and absence of the spin label or more simply by the comparison of peak intensities in HSQC spectra collected in the presence or absence of the spin label. This method was used to identify long-range interactions in the denatured state of the Δ131Δ fragment of staphylococcal nuclease labelled with the PROXYL spin label, indicating that this species is not a random coil (Gillespie and Shortle 1997). A similar study at the pH=2.3 of the acid-unfolded state of apomyoglobin labeled with the methanethiosulphonate spin label has identified the presence of transient native-like contacts (Lietzow, Jamin et al. 2002).

About a decade ago a new solution-based NMR technique that can provide atomic resolution studies of millisecond time scale conformational transitions in proteins, named Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR was developed (Palmer, Kroenke et al. 2001). Advances in the development of ^{1}H, ^{15}N, and ^{13}C NMR relaxation dispersion experiments that probe millisecond conformational exchange at nearly every site in labeled proteins have recently allowed detailed quantitative studies of processes involving multiple low-populated states (Grey, Wang et al. 2003; Boehr, McElheny et al. 2006). In cases where exchange is between a ground state and intermediate states that are populated at levels of 0.5% or higher with rates of exchange on the order of a few hundred to several thousand per second, the relaxation data measured for multiple nuclei often allow complete kinetic and thermodynamic characterization of the exchange process, even

though the resonances of the intermediate states cannot be directly observed in NMR spectra. Additionally, relaxation dispersion data report NMR chemical shift differences between states, providing structural information on the low-populated intermediate species (Korzhnev, Salvatella et al. 2004).

Neuropeptide Y family

The neuropeptide Y (NPY), the peptide YY (PYY) and the pancreatic peptide (PP) are peptides from the NPY family of neurohormones (Fig. 14) (Larhammar 1996), which are involved in the regulation of various important biological functions such as regulation of blood pressure, memory retention and food uptake to name just a few of them (Colmers and Wahlestedt 1993). These hormones target the Y receptors, G-protein coupled receptors encountered in the central and peripheral nervous system as well as in the intestinal tract (Larhammar 1996; Michel, Beck-Sickinger et al. 1998).

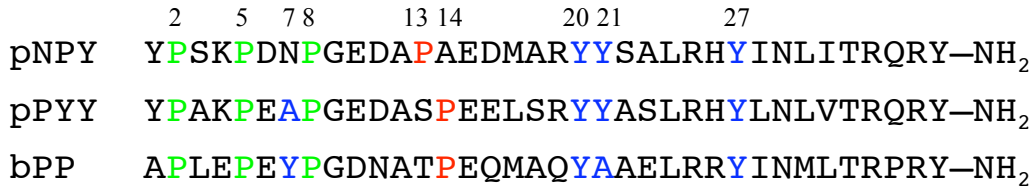


Figure 14 Sequences of NPY, PYY and bPP.

The structural features of PP were first described by Blundell et al. (Blundell, Pitts et al. 1981) in their solid-state structure of avian PP (Fig. 15), which presented one of the first crystal structures of a small peptide that lacks any disulfide bonds or other constraints. Therein, an α-helix encompassing residues 14 to 31 is preceded by a polyproline helix formed by residues 1 to 8 linked via a turn (residues 9 to 13) such that a hydrophobic contact between the C-terminal and the polyproline helices is made resulting in a hairpin-like structure. The structure of hNPY in solution was determined by Monks et al. (Monks, Karagianis et al. 1996) and displays a C-terminal α-helix comprising residues 11 to 36 (Fig. 16). The N terminus is unstructured and freely diffuses in solution as shown by the absence of medium- or long-range NOEs in that segment. In addition, the flexibility of the N terminus was verified from order parameters derived from ^{15}N relaxation data (Bader, Bettio et al. 2001) in a Lipari-

Szabo type analysis. The structure of pPYY very much resembles the fold of PP in that the N terminus folds back onto the C terminal α-helix comprising residues 17 to 31 (Fig. 15) (Keire, Kobayashi et al. 2000; Lerch, Mayrhofer et al. 2004).

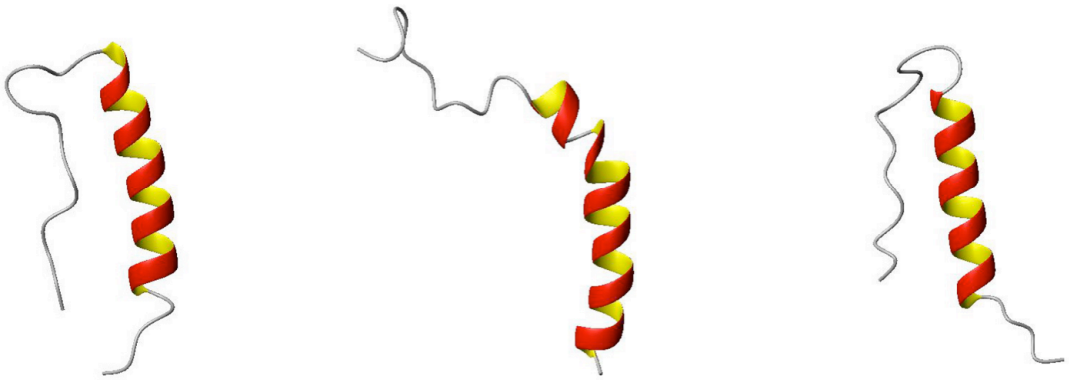


Figure 15 Structures of PYY (left), NPY(middle) and bPP (right) in water.

In addition, the crystal structure of aPP revealed that the molecule forms a symmetric dimer, in which the interface is formed by the three aromatic residues Tyr7, Phe20 and Tyr21, that form a hydrophobic cluster of aromatic residues, in which the π-systems stack onto each other. pNPY was known to exists in dimeric form and a covalently incorporated spin-label was used to establish that pNPY forms a dimer in which the α-helices from the different protomers are aligned in both parallel and antiparallel fashion (Bader, Bettio et al. 2001). In case of pPYY, in which Ala occupies position 7 instead of Tyr, the peptide is proposed to predominantly exists as a monomer at millimolar concentrations (Keire, Bowers et al. 2002). PYY presents a very interesting system that posses both secondary and tertiary, α-hairpin type, structure, a remarkable feature of such a small peptide. Studies into folding in such a small system shine light onto the mechanism of folding for globular proteins.

Our group has previously proposed that these peptides associate with the membrane (Bader and Zerbe 2005) prior to binding to the Y receptors. It was suggested that similar receptor binding profiles should be reflected in similar conformations in the state from which the peptides are recognized, and hence we have recently determined the structure of various peptides from the NPY family in order to investigate whether binding to the membrane is likely to precede receptor binding; amongst them are the neuropeptide Y (NPY) (Bader, Bettio et al. 2001), the pancreatic peptide (PP) (Lerch, Gafner et al. 2002) and the peptide YY (PYY) (Lerch,

Mayrhofer et al. 2004) in the presence of phospholipid DPC micelles serving as a membrane mimetic. Interestingly, larger structural changes occurring upon micelle binding were observed. The two peptides NPY and PYY displayed a different fold in solution whereas their conformation in the micelle-bound state is almost identical (Lerch, Mayrhofer et al. 2004). Considering their highly similar pharmacology at the Y receptors it was proposed that this is an indication that the peptides were recognized by their receptors from the micelle-bound state. The structural difference between NPY and PYY in solution is remarkable considering that both peptides are binding with highly similar affinities to the Y receptor subtypes. Moreover, the sequence homology with more than 80% is rather high and the structures therefore were also expected to be very similar.

1.4. Scope of this work

The scope of this work covers a broad range of modern solution NMR applications including investigation of protein folding processes in solution, structural studies of proteins in solution, of membrane proteins and the observation of protein dynamics in different environments. In particular the challenge of understanding the integral membrane proteins GPCRs structure and ligand interactions is one of the most exciting in the field of modern NMR. G-protein coupled receptors are of prime importance, because of their significance in biological systems and lack of structural information. During the work important insights were obtained on the GPCR fragments structure and micelle incorporation mode and on the GPCR ligands solution structure and folding.

Helical hairpin folding

In this part of the work the helical hairpin formation between a polyproline type-II helix and an α-helix as present in the natural mini-protein peptide YY (PYY) was probed. The research was carried out using site-directed mutagenesis techniques, expressing the desired ^{15}N-labeled mutant peptides in the form of a hydrophobic fusion protein followed by cleavage and purification. To reveal which residues were

responsible for the hairpin formation and how the hairpin folding occured in solution we systematically mutated Tyr and Pro residues at the hairpin interface so as in the turn region (Fig. 16).

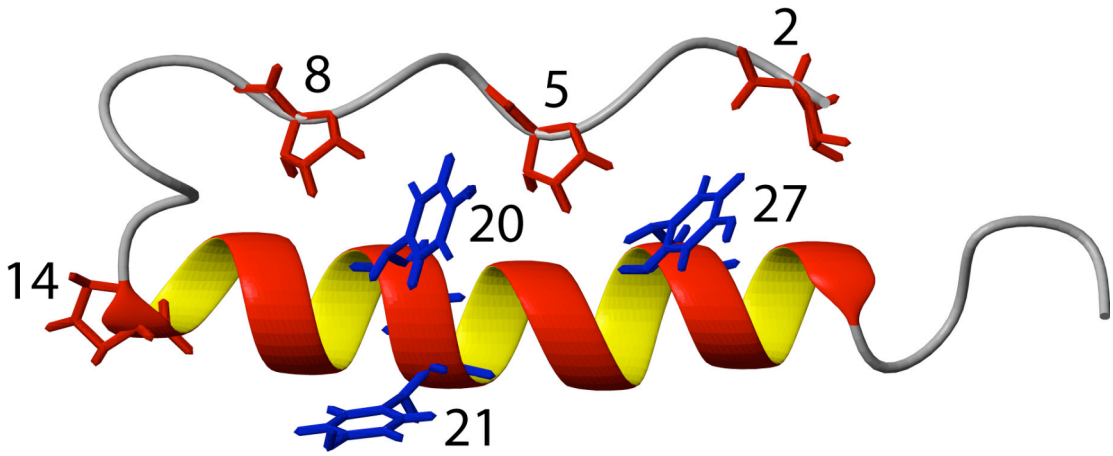


Figure 16 Structure of porcine PYY with the side-chains of Pro and Tyr residues depicted in red or blue, respectively.

By NMR spectroscopy and denaturation experiments on a series of mutants we could demonstrate that tertiary structure formation in PYY is highly dependent on the presence of a number of specific contacts. Both turn sequence and interactions of tyrosine side-chains from the C-terminal α-helix with the pockets formed by N-terminal Pro residues (Fig. 16) were shown to be important determinants of backbone dynamics and hairpin stability, suggesting a close analogy with some β-hairpin structures. It was shown that multiple relatively weak contacts between the helices were necessary for the formation of the helical hairpin, while the hairpin turn acts like a hinge, which through certain single amino acid substitutions is destabilized such that hairpin formation was completely abolished. Denaturation and renaturation of tertiary structure by temperature or co-solvents were probed by measuring changes of chemical shifts. Folding of PYY is both reversible and cooperative as inferred from the sigmoidal denaturation curves displayed by residues at the interface of the helical hairpin (*vide infra*) (Neumoin, Mares et al. 2007). The solvent mixture experiments may also simulate the structural transition occurring when a peptide diffuses from the bulk solution towards the membrane.

Structural studies of GPCR fragments

The second part of the work is devoted to the study of structure and dynamics of separated fragments of the G-protein coupled α-factor receptor Ste2p from yeast in a membrane-mimicking environment. For structural characterization we have chosen two peptides consisting of the third extracellular loop, the 7th transmembrane helix and 40 residues from the cytosolic C-terminal domain of STE2p (TM7 peptide) and another one consisting of the 19 residues from N-terminal domain, the 1st transmembrane helix, the first cytoplasmic loop, the second transmembrane helix and 7 residues from the first endoplasmic loop (TM1TM2 peptide). The alignment of the potential transmembrane domains of the peptides under investigation with those known of rhodopsin (Rho) and β2-adrenergic receptor (β2AR) is shown on Fig. 17. The TM1TM2 peptide contains a large number of polar residues in both first and second transmembrane domains compare to Rho and β2AR that gives a strong evidence of higher conformational flexibility. Furthermore the TM1TM2 sequence contains two charged residues: Arg58 residue close to the center of the first putative TM domain and His94 in the center of the second putative transmembrane helix, which are highly unfavorable when located in the center of the lipid bilayer. Hence possible difficulties could appear applying solution NMR to this system due to internal flexibility of the system and probable conformational exchange. On the other hand the TM7 peptide contains almost an equally polar putative transmembrane domain and no charged residues in comparison to Rho and β2AR, but a Pro300 residue in the middle of the putative transmembrane helix that might induce a kink in the structure.

TM1TM2

```
35-WQFSMLAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTS-98  Rho
33-VVGMGIVMSLIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVMGLAVVPFGAAHILMK-97 β2AR
49-VTQAIMFGVRCGAAALTLIVMWMTSRSRKTPIFIINQVSLFLIILHSALYFKYLLS-104         STE2p
```

TM7

```
285-PIFMTIPAFFAKTSAVYNPVIYIMMN-310    Rho
304-RKEVYILLNWIGYVNSGFNPLIYCRSP-330   β2AR
274-TDVLTTVATLLAVLSLPLSSMWATAAN-310   STE2p
```

Figure 17 Sequence alignment of the TM1TM2 and TM7 peptides of STE2p with rhodopsin and β2-adrenergic receptor fragments (transmembrane helix regions are highlighted, polar residues are marked bold, charged residues are additionally marked italic). The sequences were aligned starting from the first residue of the (predicted) membrane-spanning fragments.

We investigated the 73-residue peptide TM7 in dodecylphosphocholine (DPC) micelles. The structure revealed the presence of an α-helix in the segment of residues 10 to 30, which was perturbed around the internal Pro24 residue and displayed much flexibility around this site (Fig. 18) (Neumoin, Arshava et al. 2007).

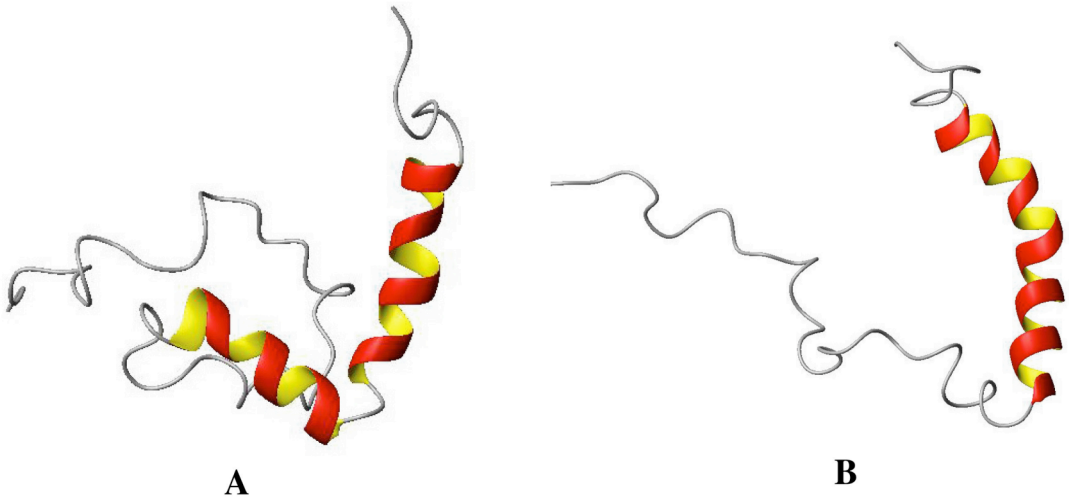


Figure 18 Individual conformers representing the structure of the TM7 peptide with flexibility around the internal Pro24 residue. A – some of the conformers calculated display a disturbed helix. B – in some of the conformers the helix is not disturbed, but kinked as well.

Solution NMR methods were used to determine the structure and topology of micelle incorporation mode of the TM7 peptide utilizing membrane-integrated (5- and 16-doxyl-stearate) and water-soluble (Gd-DOTA) spin-labels. Additional structural information was obtained from heteronuclear NOEs, residual dipolar couplings in polyacrylamide gels and hydrogen-deuterium exchange experiments.

Recently the resonance assignment of the larger 80-residue peptide TM1TM2 was accomplished on a sample in lyso-palmitoylphosphatidylglycerol (LPPG) micelles, and a structure was calculated based on the NMR data. The preliminary

structure is presented in Fig. 19, and reveals the presence of two TM helices corresponding to the putative TM1 and TM2 regions of the receptor. ^{15}N-relaxation and hydrogen-deuterium exchange data support a stable fold for the transmembrane parts of TM1TM2, whereas the solvent-exposed segments are more flexible. The exact interhelical orientation could not be determined due to the lack of a sufficient number of interhelical NOEs. The overall amount of the observed helix-helix contacts is only four, concentrated mostly close to the loop region of the peptide. Further investigations are ongoing.

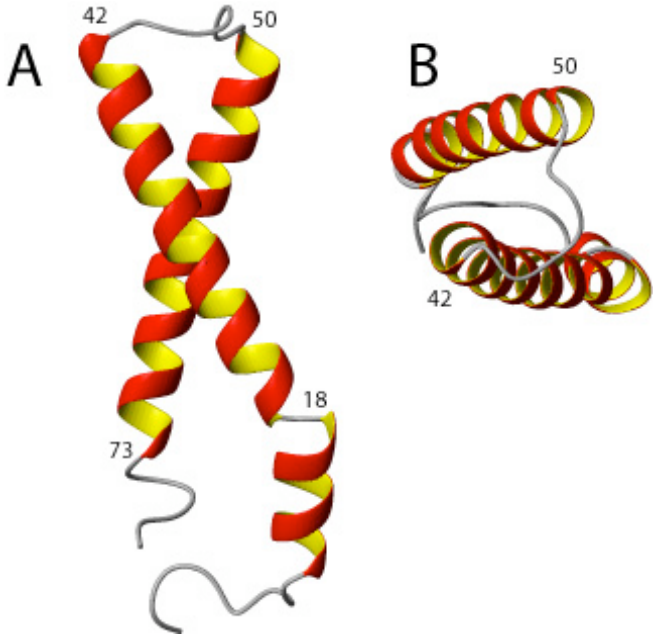


Figure 19 Structure of a single conformer of the TM1TM2 peptide: A - view from the side of membrane interior (extracellular side down, cytosolic side up); B - view from the cytoplasmic side. Residue numbers correspond to the termini of the transmembrane helices.

The present study demonstrated that polypeptides corresponding to fragments of GPCRs can be incorporated into phospholipid micelles, provided that certain protocols for incorporation are followed, and that detailed information concerning the structure of the peptide and the topology of various regions in the micelle can be extracted. Future studies will be devoted to extended systems that allow the study of larger receptor fragments and their interactions with the ligands. The present work provides an important starting point for such investigations.

1.5. References

Altenbach, C., K. Cai, et al. (1999). "Structural features and light-dependent changes in the sequence 306-322 extending from helix VII to the palmitoylation sites in rhodopsin: a site-directed spin-labeling study." Biochemistry 38(25): 7931-7.

Altenbach, C., K. Cai, et al. (2001). "Structure and function in rhodopsin: mapping light-dependent changes in distance between residue 65 in helix TM1 and residues in the sequence 306-319 at the cytoplasmic end of helix TM7 and in helix H8." Biochemistry 40(51): 15483-92.

Altenbach, C., J. Klein-Seetharaman, et al. (2001). "Structure and function in rhodopsin: mapping light-dependent changes in distance between residue 316 in helix 8 and residues in the sequence 60-75, covering the cytoplasmic end of helices TM1 and TM2 and their connection loop CL1." Biochemistry 40(51): 15493-500.

Andronesi, O. C., S. Becker, et al. (2005). "Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy." J. Am. Chem. Soc. 127(37): 12965-74.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science 181(96): 223-30.

Arai, M. and K. Kuwajima (2000). "Role of the molten globule state in protein folding." Adv. Protein Chem. 53: 209-82.

Arkin, I. T. and A. T. Brunger (1998). "Statistical analysis of predicted transmembrane alpha-helices." Biochim. Biophys. Acta 1429(1): 113-28.

Arshava, B., I. Taran, et al. (2002). "High resolution NMR analysis of the seven transmembrane domains of a heptahelical receptor in organic-aqueous medium." Biopolymers 64(3): 161-76.

Azzi, M., P. G. Charest, et al. (2003). "Beta-arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors." Proc. Natl. Acad. Sci. U S A 100(20): 11406-11.

Bader, R., A. Bettio, et al. (2001). "Structure and dynamics of micelle-bound neuropeptide Y: comparison with unligated NPY and implications for receptor selection." J. Mol. Biol. 305(2): 307-29.

Bader, R. and O. Zerbe (2005). "Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state?" Chembiochem 6(9): 1520-34.

Bai, Y., T. R. Sosnick, et al. (1995). "Protein folding intermediates: native-state hydrogen exchange." Science 269(5221): 192-7.

Barbieri, R., I. Bertini, et al. (2002). "Paramagnetically induced residual dipolar couplings for solution structure determination of lanthanide binding proteins." J. Am. Chem. Soc. 124(19): 5581-7.

Bastiaanse, E. M., K. M. Hold, et al. (1997). "The effect of membrane cholesterol content on ion transport processes in plasma membranes." Cardiovasc. Res. 33(2): 272-83.

Battiste, J. L. and G. Wagner (2000). "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data." Biochemistry 39(18): 5355-65.

Bissantz, C. (2003). "Conformational changes of G protein-coupled receptors during their activation by agonist binding." J. Recept. Signal. Transduct. Res. 23(2-3): 123-53.

Bloom, M., E. Evans, et al. (1991). "Physical properties of the fluid lipid-bilayer component of cell membranes: a perspective." Q. Rev. Biophys. 24(3): 293-397.

Blundell, T. L., J. E. Pitts, et al. (1981). "X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone." Proc. Natl. Acad. Sci. U S A 78(7): 4175-4179.

Boehr, D. D., D. McElheny, et al. (2006). "The dynamic energy landscape of dihydrofolate reductase catalysis." Science 313(5793): 1638-42.

Borhan, B., M. L. Souto, et al. (2000). "Movement of retinal along the visual transduction path." Science 288(5474): 2209-12.

Bouvier, M. (2001). "Oligomerization of G-protein-coupled transmitter receptors." Nat. Rev. Neurosci. 2(4): 274-86.

Breitwieser, G. E. (2004). "G protein-coupled receptor oligomerization: implications for G protein activation and cell signaling." Circ. Res. 94(1): 17-27.

Brown, L. R. and K. Wuthrich (1981). "Melittin bound to dodecylphosphocholine micelles. H-NMR assignments and global conformational features." Biochim. Biophys. Acta 647(1): 95-111.

Call, M. E., J. R. Schnell, et al. (2006). "The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor." Cell 127(2): 355-68.

Calver, A. R., M. J. Robbins, et al. (2001). "The C-terminal domains of the GABA(b) receptor subunits mediate intracellular trafficking but are not required for receptor signaling." J. Neurosci. 21(4): 1203-10.

Cardoso, J. C., V. C. Pinto, et al. (2006). "Evolution of secretin family GPCR members in the metazoa." BMC Evol. Biol. 6: 108.

Carroll, F. Y., A. Stolle, et al. (2001). "BAY36-7620: a potent non-competitive mGlu1 receptor antagonist with inverse agonist activity." Mol. Pharmacol. 59(5): 965-73.

Chabre, M. and M. le Maire (2005). "Monomeric G-protein-coupled receptor as a functional unit." Biochemistry 44(27): 9395-403.

Chamberlain, A. K., Y. Lee, et al. (2004). "Snorkeling preferences foster an amino acid composition bias in transmembrane helices." J. Mol. Biol. 339(2): 471-9.

Cheng, B. Y., J. G. Carbonell, et al. (2005). "Protein classification based on text document classification techniques." Proteins 58(4): 955-70.

Cherezov, V., D. M. Rosenbaum, et al. (2007). "High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor." Science 318(5854): 1258-65.

Choi, G., J. Landin, et al. (2002). "Structural studies of metarhodopsin II, the activated form of the G-protein coupled receptor, rhodopsin." Biochemistry 41(23): 7318-24.

Chou, J. J., S. Gaemers, et al. (2001). "A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles." J. Biomol. NMR 21(4): 377-82.

Chou, J. J., J. D. Kaufman, et al. (2002). "Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamide gel." J. Am. Chem. Soc. 124(11): 2450-1.

Christopoulos, A. and T. Kenakin (2002). "G protein-coupled receptor allosterism and complexing." Pharmacol. Rev. 54(2): 323-74.

Cierpicki, T. and J. H. Bushweller (2004). "Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins." J. Am. Chem. Soc. 126(49): 16259-66.

Cierpicki, T., B. Liang, et al. (2006). "Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A." J. Am. Chem. Soc. 128(21): 6947-51.

Colmers, W. and C. Wahlestedt (1993). The Biology of Neuropeptide Y and Related Peptides, Humana Press.

Cordes, F. S., J. N. Bright, et al. (2002). "Proline-induced distortions of transmembrane helices." J. Mol. Biol. 323(5): 951-60.

Cornilescu, G., F. Delaglio, et al. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J. Biomol. NMR 13(3): 289-302.

Cowan, S. W., T. Schirmer, et al. (1992). "Crystal structures explain functional properties of two E. coli porins." Nature 358(6389): 727-33.

Crocker, E., M. Eilers, et al. (2006). "Location of Trp265 in metarhodopsin II: implications for the activation mechanism of the visual receptor rhodopsin." J. Mol. Biol. 357(1): 163-72.

Das, S. S. and G. A. Banker (2006). "The role of protein interaction motifs in regulating the polarity and clustering of the metabotropic glutamate receptor mGluR1a." J. Neurosci. 26(31): 8115-25.

de Planque, M. R., B. B. Bonev, et al. (2003). "Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions." Biochemistry 42(18): 5341-8.

de Planque, M. R., E. Goormaghtigh, et al. (2001). "Sensitivity of single membrane-spanning alpha-helical peptides to hydrophobic mismatch with a lipid bilayer: effects on backbone structure, orientation, and extent of membrane incorporation." Biochemistry 40(16): 5000-10.

de Planque, M. R., D. V. Greathouse, et al. (1998). "Influence of lipid/peptide hydrophobic mismatch on the thickness of diacylphosphatidylcholine bilayers. A 2H NMR and ESR study using designed transmembrane alpha-helical peptides and gramicidin A." Biochemistry 37(26): 9333-45.

de Planque, M. R. and J. A. Killian (2003). "Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring." Mol. Membr. Biol. 20(4): 271-84.

Dean, M. K., C. Higgs, et al. (2001). "Dimerization of G-protein-coupled receptors." J. Med. Chem. 44(26): 4595-614.

Delaglio, F., G. Kontaxis, et al. (2000). "Protein Structure Determination Using Molecular Fragment Replacement and NMR Dipolar Couplings." J. Am. Chem. Soc. 122(9): 2142-2143.

Demene, H., S. Granier, et al. (2003). "Active peptidic mimics of the second intracellular loop of the V(1A) vasopressin receptor are structurally related to the second intracellular rhodopsin loop: a combined 1H NMR and biochemical study." Biochemistry 42(27): 8204-13.

Devaux, P. F. (1993). "Lipid transmembrane asymmetry and flip-flop in biological membranes and in lipid bilayers" Curr. Opin. Struct. Biol. 3(4): 489-494.

Douglas, S. M., J. J. Chou, et al. (2007). "DNA-nanotube-induced alignment of membrane proteins for NMR structure determination." Proc. Natl. Acad. Sci. U S A 104(16): 6644-8.

Dowhan, W. (1997). "Molecular basis for membrane phospholipid diversity: Why are there so many lipids?" Annu. Rev. Biochem. 66: 199-232.

Dvoretsky, A., V. Gaponenko, et al. (2002). "Derivation of structural restraints using a thiol-reactive chelator." FEBS Lett. 528(1-3): 189-92.

Dyson, H. J., M. Rance, et al. (1988). "Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn." J. Mol. Biol. 201(1): 161-200.

Eilers, M., S. C. Shekar, et al. (2000). "Internal packing of helical membrane proteins." Proc. Natl. Acad. Sci. U S A 97(11): 5796-801.

Eliezer, D., J. Chung, et al. (2000). "Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin." Biochemistry 39(11): 2894-901.

Estephan, R., J. Englander, et al. (2005). "Biosynthesis and NMR analysis of a 73-residue domain of a Saccharomyces cerevisiae G protein-coupled receptor." Biochemistry 44(35): 11795-810.

Farrens, D. L., C. Altenbach, et al. (1996). "Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin." Science 274(5288): 768-70.

Filipek, S., K. A. Krzysko, et al. (2004). "A concept for G protein activation by G protein-coupled receptor dimers: the transducin/rhodopsin interface." Photochem. Photobiol. Sci. 3(6): 628-38.

Filizola, M. and H. Weinstein (2002). "Structural models for dimerization of G-protein coupled receptors: the opioid receptor homodimers." Biopolymers 66(5): 317-25.

Fotiadis, D., Y. Liang, et al. (2003). "Atomic-force microscopy: Rhodopsin dimers in native disc membranes." Nature 421(6919): 127-8.

Fotiadis, D., Y. Liang, et al. (2004). "The G protein-coupled receptor rhodopsin in the native membrane." FEBS Lett. 564(3): 281-8.

Fridmanis, D., R. Fredriksson, et al. (2007). "Formation of new genes explains lower intron density in mammalian Rhodopsin G protein-coupled receptors." Mol. Phylogenet. Evol. 43(3): 864-80.

Gennis, R. B. (1989). Biomembranes: Molecular Structure and Function. New York, Springer.

George, S. R., B. F. O'Dowd, et al. (2002). "G-protein-coupled receptor oligomerization and its potential for drug discovery." Nat. Rev. Drug. Discov. 1(10): 808-20.

Gether, U. (2000). "Uncovering molecular mechanisms involved in activation of G protein-coupled receptors." Endocr. Rev. 21(1): 90-113.

Gether, U., F. Asmar, et al. (2002). "Structural basis for activation of G-protein-coupled receptors." Pharmacol. Toxicol. 91(6): 304-12.

Gillespie, J. R. and D. Shortle (1997). "Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures." J. Mol. Biol. 268(1): 170-84.

Girvin, M. E., V. K. Rastogi, et al. (1998). "Solution structure of the transmembrane H+-transporting subunit c of the F1F0 ATP synthase." Biochemistry 37(25): 8817-24.

Gloriam, D. E., H. B. Schioth, et al. (2005). "Nine new human Rhodopsin family G-protein coupled receptors: identification, sequence characterisation and evolutionary relationship." Biochim. Biophys. Acta 1722(3): 235-46.

Granseth, E., G. von Heijne, et al. (2005). "A study of the membrane-water interface region of membrane proteins." J. Mol. Biol. 346(1): 377-85.

Gray, T. M. and B. W. Matthews (1984). "Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins." J. Mol. Biol. 175(1): 75-81.

Grey, M. J., C. Wang, et al. (2003). "Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling." J. Am. Chem. Soc. 125(47): 14324-35.

Griffiths, J. M., K. V. Lakshmi, et al. (1994). "Dipolar Correlation NMR Spectroscopy of a Membrane Protein." J. Am. Chem. Soc. 116(22): 10178-10181.

Guo, W., L. Shi, et al. (2003). "The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer." J. Biol. Chem. 278(7): 4385-8.

Harzer, U. and B. Bechinger (2000). "Alignment of lysine-anchored membrane peptides under conditions of hydrophobic mismatch: a CD, 15N and 31P solid-state NMR spectroscopy investigation." Biochemistry 39(43): 13106-14.

Hebert, T. E. and M. Bouvier (1998). "Structural and functional aspects of G protein-coupled receptor oligomerization." Biochem. Cell Biol. 76(1): 1-11.

Hebert, T. E., S. Moffett, et al. (1996). "A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation." J. Biol. Chem. 271(27): 16384-92.

Henderson, R., J. M. Baldwin, et al. (1990). "Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy." J. Mol. Biol. 213(4): 899-929.

Henry, G. D. and B. D. Sykes (1994). "Methods to study membrane protein structure in solution." Methods Enzymol. 239: 515-35.

Herzfeld, J. (1996). "Entropically Driven Order in Crowded Solutions: From Liquid Crystals to Cell Biology." Acc. Chem. Res. 29(1): 31-37.

Hink, M. A., A. van Hoek, et al. (1999). "Dynamics of Phospholipid Molecules in Micelles: Characterization with Fluorescence Correlation Spectroscopy and Time-Resolved Fluorescence Anisotropy." Langmuir 15(4): 992-997.

Horn, F., E. Bettler, et al. (2003). "GPCRDB information system for G protein-coupled receptors." Nucleic Acids Res. 31(1): 294-7.

Hwang, P. M. and L. E. Kay (2005). "Solution structure and dynamics of integral membrane proteins by NMR: a case study involving the enzyme PagP." Methods Enzymol. 394: 335-50.

Ikegami, T., L. Verdier, et al. (2004). "Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions." J. Biomol. NMR 29(3): 339-49.

Jackson, S. E. (1998). "How do small single-domain proteins fold?" Fold. Des. 3(4): R81-91.

Jackson, S. E. and A. R. Fersht (1991). "Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition." Biochemistry 30(43): 10428-35.

Ji, T. H., M. Grossmann, et al. (1998). "G protein-coupled receptors. I. Diversity of receptor-ligand interactions." J. Biol. Chem. 273(28): 17299-302.

John Cavanagh, W. J. F., Arthur G. Palmer III, Nicholas J. Skelton (1995). Protein NMR Spectroscopy: Principles and Practice, Academic Press.

Jordan, B. A. and L. A. Devi (1999). "G-protein-coupled receptor heterodimerization modulates receptor function." Nature 399(6737): 697-700.

Kaiser, E. T. and F. J. Kezdy (1983). "Secondary structures of proteins and peptides in amphiphilic environments. (A review)." Proc. Natl. Acad. Sci. U S A 80(4): 1137-43.

Kamen, D. E., S. M. Cahill, et al. (2007). "Multiple alignment of membrane proteins for measuring residual dipolar couplings using lanthanide ions bound to a small metal chelator." J. Am. Chem. Soc. 129(7): 1846-7.

Kandasamy, S. K. and R. G. Larson (2006). "Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch." Biophys. J. 90(7): 2326-43.

Katragadda, M., J. L. Alderfer, et al. (2000). "Solution structure of the loops of bacteriorhodopsin closely resembles the crystal structure." Biochim. Biophys. Acta 1466(1-2): 1-6.

Katragadda, M., J. L. Alderfer, et al. (2001). "Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin." Biophys. J. 81(2): 1029-36.

Keire, D. A., C. W. Bowers, et al. (2002). "Structure and receptor binding of PYY analogs." Peptides 23(2): 305-21.

Keire, D. A., M. Kobayashi, et al. (2000). "Solution structure of monomeric peptide YY supports the functional significance of the PP-fold." Biochemistry 39(32): 9935-42.

Killian, J. A. (1998). "Hydrophobic mismatch between proteins and lipids in membranes." Biochim. Biophys. Acta 1376(3): 401-15.

Kim, P. S. and R. L. Baldwin (1982). "Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding." Annu. Rev. Biochem. 51: 459-89.

Kisselev, O. G., J. Kao, et al. (1998). "Light-activated rhodopsin induces structural binding motif in G protein alpha subunit." Proc. Natl. Acad. Sci. U S A 95(8): 4270-5.

Klabunde, T. and G. Hessler (2002). "Drug design strategies for targeting G-protein-coupled receptors." Chembiochem 3(10): 928-44.

Klco, J. M., T. B. Lassere, et al. (2003). "C5a receptor oligomerization. I. Disulfide trapping reveals oligomers and potential contact surfaces in a G protein-coupled receptor." J. Biol. Chem. 278(37): 35345-53.

Koenig, B. W., D. C. Mitchell, et al. (2000). "Measurement of dipolar couplings in a transducin peptide fragment weakly bound to oriented photo-activated rhodopsin." J. Biomol. NMR 16(2): 121-5.

Kolakowski, L. F., Jr. (1994). "GCRDb: a G-protein-coupled receptor database." Receptors Channels 2(1): 1-7.

Kornberg, R. D. and H. M. McConnel (1971). "Lateral diffusion of phospholipids in a vesicle membrane." Proc. Natl. Acad. Sci. USA 68: 2564.

Korzhnev, D. M., X. Salvatella, et al. (2004). "Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR." Nature 430(6999): 586-90.

Krishna, M. M., L. Hoang, et al. (2004). "Hydrogen exchange methods to study protein folding." Methods 34(1): 51-64.

Kroeger, K. M., A. C. Hanyaloglu, et al. (2001). "Constitutive and agonist-dependent homo-oligomerization of the thyrotropin-releasing hormone receptor. Detection in living cells using bioluminescence resonance energy transfer." J. Biol. Chem. 276(16): 12736-43.

Krueger-Koplin, R. D., P. L. Sorgen, et al. (2004). "An evaluation of detergents for NMR structural studies of membrane proteins." J. Biomol. NMR 28(1): 43-57.

Ladokhin, A. S. and S. H. White (1999). "Folding of amphipathic alpha-helices on membranes: energetics of helix formation by melittin." J. Mol. Biol. 285(4): 1363-9.

Larhammar, D. (1996). "Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide." Regul. Pept. 65(3): 165-74.

Lazarova, T., K. A. Brewin, et al. (2004). "Characterization of peptides corresponding to the seven transmembrane domains of human adenosine A2a receptor." Biochemistry 43(40): 12945-54.

Lee, A. G. (2004). "How lipids affect the activities of integral membrane proteins." Biochim. Biophys. Acta 1666(1-2): 62-87.

Lefkowitz, R. J. and S. K. Shenoy (2005). "Transduction of receptor signals by beta-arrestins." Science 308(5721): 512-7.

Lerch, M., V. Gafner, et al. (2002). "Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles." J. Mol. Biol. 322(5): 1117-33.

Lerch, M., M. Mayrhofer, et al. (2004). "Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors." J. Mol. Biol. 339(5): 1153-68.

Levinthal, C. (1968). "Are there pathways for protein folding?" J. Chim. Phys. 65: 44-45.

Lewis, B. A. and D. M. Engelman (1983). "Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles." J. Mol. Biol. 166(2): 211-7.

Liang, B., J. H. Bushweller, et al. (2006). "Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy." J. Am. Chem. Soc. 128(13): 4389-97.

Lichtenberg, D., R. J. Robson, et al. (1983). "Solubilization of phospholipids by detergents. Structural and kinetic aspects." Biochim. Biophys. Acta 737(2): 285-304.

Lietzow, M. A., M. Jamin, et al. (2002). "Mapping long-range contacts in a highly unfolded protein." J. Mol. Biol. 322(4): 655-62.

Lu, H., T. Marti, et al. (2001). "Proline residues in transmembrane alpha helices affect the folding of bacteriorhodopsin." J. Mol. Biol. 308(2): 437-46.

Luca, S., H. Heise, et al. (2003). "High-resolution solid-state NMR applied to polypeptides and membrane proteins." Acc. Chem. Res. 36(11): 858-65.

Luttrell, L. M. and R. J. Lefkowitz (2002). "The role of beta-arrestins in the termination and transduction of G-protein-coupled receptor signals." J. Cell. Sci. 115(Pt 3): 455-65.

Ma, C. and S. J. Opella (2000). "Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy." J. Magn. Reson. 146(2): 381-4.

MacKenzie, K. R., J. H. Prestegard, et al. (1997). "A transmembrane helix dimer: structure and implications." Science 276(5309): 131-3.

Maggio, R., Z. Vogel, et al. (1993). "Coexpression studies with mutant muscarinic/adrenergic receptors provide evidence for intermolecular "cross-talk" between G-protein-linked receptors." Proc. Natl. Acad. Sci. U S A 90(7): 3103-7.

McIntosh, T. J. and S. A. Simon (2006). "Roles of bilayer material properties in function and distribution of membrane proteins." Annu. Rev. Biophys. Biomol. Struct. 35: 177-98.

McLean, L. R., S. H. Buck, et al. (1990). "Examination of the role of the amphipathic alpha-helix in the interaction of neuropeptide Y and active cyclic analogues with cell membrane receptors and dimyristoylphosphatidylcholine." Biochemistry 29(8): 2016-22.

Meng, E. C. and H. R. Bourne (2001). "Receptor activation: what does the rhodopsin structure tell us?" Trends. Pharmacol. Sci. 22(11): 587-93.

Michel, M. C., A. Beck-Sickinger, et al. (1998). "XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors." Pharmacol. Rev. 50(1): 143-50.

Minakata, H., J. W. Taylor, et al. (1989). "Characterization of amphiphilic secondary structures in neuropeptide Y through the design, synthesis, and study of model peptides." J. Biol. Chem. 264(14): 7907-13.

Miura, S., Y. H. Feng, et al. (1999). "Role of aromaticity of agonist switches of angiotensin II in the activation of the AT1 receptor." J. Biol. Chem. 274(11): 7103-10.

Moller, J. V. and M. le Maire (1993). "Detergent binding as a measure of hydrophobic surface area of integral membrane proteins." J. Biol. Chem. 268(25): 18659-72.

Monks, S. A., G. Karagianis, et al. (1996). "Solution structure of human neuropeptide Y." J. Biomol. NMR 8(4): 379-90.

Moroder, L., R. Romano, et al. (1993). "New evidence for a membrane-bound pathway in hormone receptor binding." Biochemistry 32(49): 13551-9.

Mouritsen, O. G. and M. Bloom (1984). "Mattress model of lipid-protein interactions in membranes." Biophys. J. 46(2): 141-53.

Munoz, V. and J. M. Sanchez-Ruiz (2004). "Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments." Proc. Natl. Acad. Sci. U S A 101(51): 17646-51.

Munro, S. (1995). "An investigation of the role of transmembrane domains in Golgi protein retention." EMBO J. 14(19): 4695-704.

Munro, S. (2003). "Lipid rafts: elusive or illusive?" Cell 115(4): 377-88.

Nakagawa, T., T. Sakurai, et al. (2005). "Insect sex-pheromone signals mediated by specific combinations of olfactory receptors." Science 307(5715): 1638-42.

Neugebauer, J. (1988). A Guide to the Properties and Uses of Detergents in Biology and Biochemistry. La Jolla, CA, Calbiochem Biochemicals.

Neumoin, A., B. Arshava, et al. (2007). "NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor from Saccharomyces cerevisiae." Biophys. J. 93(2): 467-82.

Neumoin, A., J. Mares, et al. (2007). "Probing the formation of stable tertiary structure in a model miniprotein at atomic resolution: determinants of stability of a helical hairpin." J. Am. Chem. Soc. 129(28): 8811-7.

Nickell, S., P. S. Park, et al. (2007). "Three-dimensional architecture of murine rod outer segments determined by cryoelectron tomography." J. Cell. Biol. 177(5): 917-25.

Opella, S. J. and F. M. Marassi (2004). "Structure determination of membrane proteins by NMR spectroscopy." Chem. Rev. 104(8): 3587-606.

Overton, M. C. and K. J. Blumer (2002). "The extracellular N-terminal domain and transmembrane domains 1 and 2 mediate oligomerization of a yeast G protein-coupled receptor." J. Biol. Chem. 277(44): 41463-72.

Oxenoid, K. and J. J. Chou (2005). "The structure of phospholamban pentamer reveals a channel-like architecture in membranes." Proc. Natl. Acad. Sci. U S A 102(31): 10870-5.

Ozdirekcan, S., D. T. Rijkers, et al. (2005). "Influence of flanking residues on tilt and rotation angles of transmembrane peptides in lipid bilayers. A solid-state 2H NMR study." Biochemistry 44(3): 1004-12.

Pace, A. J., L. Gama, et al. (1999). "Dimerization of the calcium-sensing receptor occurs within the extracellular domain and is eliminated by Cys --> Ser mutations at Cys101 and Cys236." J. Biol. Chem. 274(17): 11629-34.

Palczewski, K., T. Kumasaka, et al. (2000). "Crystal structure of rhodopsin: A G protein-coupled receptor." Science 289(5480): 739-45.

Palmer, A. G., 3rd, C. D. Kroenke, et al. (2001). "Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules." Methods Enzymol. 339: 204-38.

Papavoine, C. H., J. M. Aelen, et al. (1995). "NMR studies of the major coat protein of bacteriophage M13. Structural information of gVIIIp in dodecylphosphocholine micelles." Eur. J. Biochem. 232(2): 490-500.

Park, P. S., S. Filipek, et al. (2004). "Oligomerization of G protein-coupled receptors: past, present, and future." Biochemistry 43(50): 15643-56.

Park, P. S., D. T. Lodowski, et al. (2008). "Activation of g protein-coupled receptors: beyond two-state models and tertiary conformational changes." Annu. Rev. Pharmacol. Toxicol. 48: 107-41.

Park, S. H., A. A. Mrse, et al. (2006). "Rotational diffusion of membrane proteins in aligned phospholipid bilayers by solid-state NMR spectroscopy." J. Magn. Reson. 178(1): 162-5.

Park, S. H., S. Prytulla, et al. (2006). "High-resolution NMR spectroscopy of a GPCR in aligned bicelles." J. Am. Chem. Soc. 128(23): 7402-3.

Parnot, C. and B. Kobilka (2004). "Toward understanding GPCR dimers." Nat. Struct. Mol. Biol. 11(8): 691-2.

Parton, R. G. and A. A. Richards (2003). "Lipid rafts and caveolae as portals for endocytosis: new insights and common mechanisms." Traffic 4(11): 724-38.

Patel, A. B., E. Crocker, et al. (2004). "Coupling of retinal isomerization to the activation of rhodopsin." Proc. Natl. Acad. Sci. U S A 101(27): 10048-53.

Pellegrini, M. and D. F. Mierke (1999). "Structural characterization of peptide hormone/receptor interactions by NMR spectroscopy." Biopolymers 51(3): 208-20.

Pin, J. P., T. Galvez, et al. (2003). "Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors." Pharmacol. Ther. 98(3): 325-54.

Popot, J. L. and D. M. Engelman (2000). "Helical membrane protein folding, stability, and evolution." Annu. Rev. Biochem. 69: 881-922.

Porter, J. E., J. Hwa, et al. (1996). "Activation of the alpha1b-adrenergic receptor is initiated by disruption of an interhelical salt bridge constraint." J. Biol. Chem. 271(45): 28318-23.

Prabhu, Y. and L. Eichinger (2006). "The Dictyostelium repertoire of seven transmembrane domain receptors." Eur. J. Cell. Biol. 85(9-10): 937-46.

Ram, P. and J. H. Prestegard (1988). "Magnetic field induced ordering of bile salt/phospholipid micelles: new media for NMR structural investigations." Biochim. Biophys. Acta 940(2): 289-94.

Rand, R. P. and V. A. Parsegian (1989). "Hydration forces between phospholipid bilayers." Biochim. Biophys. Acta 988(3): 351-376.

Rashid, A. J., B. F. O'Dowd, et al. (2004). "Minireview: Diversity and complexity of signaling through peptidergic G protein-coupled receptors." Endocrinology 145(6): 2645-52.

Rasmussen, S. G., H. J. Choi, et al. (2007). "Crystal structure of the human beta2 adrenergic G-protein-coupled receptor." Nature 450(7168): 383-7.

Rastogi, V. K. and M. E. Girvin (1999). "Structural changes linked to proton translocation by subunit c of the ATP synthase." Nature 402(6759): 263-8.

Ridder, A. N., R. E. Spelbrink, et al. (2004). "Photo-crosslinking analysis of preferential interactions between a transmembrane peptide and matching lipids." Biochemistry 43(15): 4482-9.

Rigaud, J. L., D. Levy, et al. (1998). "Detergent removal by non-polar polystyrene beads." Eur. Biophys. J. 27(4): 305-319.

Rigaud, J. L., B. Pitard, et al. (1995). "Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins." Biochim. Biophys. Acta 1231(3): 223-46.

Ruan, K. H., S. P. So, et al. (2001). "Solution structure of the second extracellular loop of human thromboxane A2 receptor." Biochemistry 40(1): 275-80.

Sadqi, M., D. Fushman, et al. (2006). "Atom-by-atom analysis of global downhill protein folding." Nature 442(7100): 317-21.

Salim, K., T. Fenton, et al. (2002). "Oligomerization of G-protein-coupled receptors shown by selective co-immunoprecipitation." J. Biol. Chem. 277(18): 15482-5.

Sanders, C. R., 2nd and G. C. Landis (1995). "Reconstitution of membrane proteins into lipid-rich bilayered mixed micelles for NMR studies." Biochemistry 34(12): 4030-40.

Sanders, C. R., 2nd and J. P. Schwonek (1993). "An approximate model and empirical energy function for solute interactions with a water-phosphatidylcholine interface." Biophys. J. 65(3): 1207-18.

Sanders, C. R., B. J. Hare, et al. (1994). "Magnetically-oriented phospholipid micelles as a tool for the study of membrane-associated molecules." Prog. Nuclear Magn. Res. Spectr. 26(Part 5): 421-444.

Sanders, C. R. and F. Sonnichsen (2006). "Solution NMR of membrane proteins: practice and challenges." Magn. Reson. Chem. 44 Spec No: S24-40.

Sankararamakrishnan, R. (2006). "Recognition of GPCRs by peptide ligands and membrane compartments theory: structural studies of endogenous peptide hormones in membrane environment." Biosci. Rep. 26(2): 131-58.

Schoneberg, T., A. Schulz, et al. (2004). "Mutant G-protein-coupled receptors as a cause of human diseases." Pharmacol. Ther. 104(3): 173-206.

Schwyzer, R. (1991). "Peptide-membrane interactions and a new principle in quantitative structure-activity relationships." Biopolymers 31(6): 785-92.

Schwyzer, R. (1995). "In search of the 'bio-active conformation'--is it induced by the target cell membrane?" J. Mol. Recognit. 8(1-2): 3-8.

Seigneuret, M. and M. Kainosho (1993). "Localisation of methionine residues in bacteriorhodopsin by carbonyl 13C-NMR with sequence-specific assignments." FEBS Lett. 327(1): 7-12.

Siegel, D. P., V. Cherezov, et al. (2006). "Transmembrane peptides stabilize inverted cubic phases in a biphasic length-dependent manner: implications for protein-induced membrane fusion." Biophys. J. 90(1): 200-11.

Silvius, J. R. (1992). "Solubilization and functional reconstitution of biomembrane components." Annu. Rev. Biophys. Biomol. Struct. 21: 323-48.

Simons, K. and W. L. Vaz (2004). "Model systems, lipid rafts, and cell membranes." Annu. Rev. Biophys. Biomol. Struct. 33: 269-95.

Singer, S. J. and G. L. Nicolson (1972). "The Fluid Mosaic Model of the Structure of Cell Membranes." Science 175: 720-731.

Smith, S. O., K. Aschheim, et al. (1996). "Magic angle spinning NMR spectroscopy of membrane proteins." Q. Rev. Biophys. 29(4): 395-449.

Sparr, E., W. L. Ash, et al. (2005). "Self-association of transmembrane alpha-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch." J. Biol. Chem. 280(47): 39324-31.

Spector, A. A. and M. A. Yorek (1985). "Membrane lipid composition and cellular function." J. Lipid. Res. 26(9): 1015-35.

Spruijt, R. B., C. J. Wolfs, et al. (2004). "Membrane assembly of M13 major coat protein: evidence for a structural adaptation in the hinge region and a tilted transmembrane domain." Biochemistry 43(44): 13972-80.

Strandberg, E. and J. A. Killian (2003). "Snorkeling of lysine side chains in transmembrane helices: how easy can it get?" FEBS Lett. 544(1-3): 69-73.

Tanford, C. (1980). The Hydrophobic Effect: Formation of Micelles and Biological Membranes. New York, John Wiley and Sons.

Teller, D. C., T. Okada, et al. (2001). "Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs)." Biochemistry 40(26): 7761-72.

Thevenin, D., T. Lazarova, et al. (2005). "Oligomerization of the fifth transmembrane domain from the adenosine A2A receptor." Protein. Sci. 14(8): 2177-86.

Thevenin, D., M. F. Roberts, et al. (2005). "Identifying interactions between transmembrane helices from the adenosine A2A receptor." Biochemistry 44(49): 16239-45.

Tian, C., R. M. Breyer, et al. (2005). "Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor." J. Am. Chem. Soc. 127(22): 8010-1.

Tian, C., R. M. Breyer, et al. (2006). "Solution NMR Spectroscopy of the Human Vasopressin V2 Receptor, A G Protein-Coupled Receptor" J. Am. Chem. Soc. 128(15): 5300-5300.

Tian, C., M. D. Karra, et al. (2005). "Membrane protein preparation for TROSY NMR screening." Methods Enzymol. 394: 321-34.

Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science 278(5340): 1111-4.

Tolman, J. R., J. M. Flanagan, et al. (1995). "Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution." Proc. Natl. Acad. Sci. U S A 92(20): 9279-83.

Tycko, R., F. J. Blanco, et al. (2000). "Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR." J. Am. Chem. Soc. 122(38): 9340-9341.

Ulfers, A. L., A. Piserchio, et al. (2002). "Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P." Biopolymers 66(5): 339-49.

Ulmschneider, M. B. and M. S. Sansom (2001). "Amino acid distributions in integral membrane protein structures." Biochim. Biophys. Acta 1512(1): 1-14.

van Duyl, B. Y., D. T. Rijkers, et al. (2002). "Influence of hydrophobic mismatch and palmitoylation on the association of transmembrane alpha-helical peptides with detergent-resistant membranes." FEBS Lett. 523(1-3): 79-84.

Veglia, G. and S. J. Opella (2000). "Lanthanide Ion Binding to Adventitious Sites Aligns Membrane Proteins in Micelles for Solution NMR Spectroscopy." J. Am. Chem. Soc. 122(47): 11733-11734.

Venturoli, M., B. Smit, et al. (2005). "Simulation studies of protein-induced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins." Biophys. J. 88(3): 1778-98.

Vidal, A. and T. J. McIntosh (2005). "Transbilayer peptide sorting between raft and nonraft bilayers: comparisons of detergent extraction and confocal microscopy." Biophys. J. 89(2): 1102-8.

Vinogradova, O., P. Badola, et al. (1997). "Escherichia coli diacylglycerol kinase: a case study in the application of solution NMR methods to an integral membrane protein." Biophys. J. 72(6): 2688-701.

Vold, R. R., R. S. Prosser, et al. (1997). "Isotropic solutions of phospholipid bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides." J. Biomol. NMR 9(3): 329-35.

Walker, J. E., A. F. Carne, et al. (1979). "The topography of the purple membrane." Nature 278(653-654).

Wallin, E., T. Tsukihara, et al. (1997). "Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria." Protein. Sci. 6(4): 808-15.

Wallin, E. and G. von Heijne (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." Protein. Sci. 7(4): 1029-38.

Watts, A., A. S. Ulrich, et al. (1995). "Membrane protein structure: the contribution and potential of novel solid state NMR approaches." Mol. Membr. Biol. 12(3): 233-46.

Weiss, T. M., P. C. van der Wel, et al. (2003). "Hydrophobic mismatch between helices and lipid bilayers." Biophys. J. 84(1): 379-85.

Wess, J. (1998). "Molecular basis of receptor/G-protein-coupling selectivity." Pharmacol. Ther. 80(3): 231-64.

White, S. H. and W. C. Wimley (1998). "Hydrophobic interactions of peptides with membrane interfaces." Biochim. Biophys. Acta 1376(3): 339-52.

White, S. H. and W. C. Wimley (1999). "Membrane protein folding and stability: physical principles." Annu. Rev. Biophys. Biomol. Struct. 28: 319-65.

Wiener, M. C. and S. H. White (1992). "Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. II. Distribution and packing of terminal methyl groups." Biophys. J. 61(2): 428-33.

Wimley, W. C. (2002). "Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures." Protein. Sci. 11(2): 301-12.

Wimley, W. C. (2003). "The versatile beta-barrel membrane protein." Curr. Opin. Struct. Biol. 13(4): 404-11.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. New York, Wiley-Interscience.

Yano, Y. and K. Matsuzaki (2006). "Measurement of thermodynamic parameters for hydrophobic mismatch 1: self-association of a transmembrane helix." Biochemistry 45(10): 3370-8.

Yau, W. M., W. C. Wimley, et al. (1998). "The preference of tryptophan for membrane interfaces." Biochemistry 37(42): 14713-8.

Yeagle, P. L. and A. D. Albert (2003). "A conformational trigger for activation of a G protein by a G protein-coupled receptor." Biochemistry 42(6): 1365-8.

Yeagle, P. L. and A. D. Albert (2007). "G-protein coupled receptor structure." Biochim. Biophys. Acta 1768(4): 808-24.

Yeagle, P. L., J. L. Alderfer, et al. (1995). "Structure of the carboxy-terminal domain of bovine rhodopsin." Nat. Struct. Biol. 2(10): 832-4.

Yeagle, P. L., J. L. Alderfer, et al. (1995). "Structure of the third cytoplasmic loop of bovine rhodopsin." Biochemistry 34(45): 14621-5.

Yeagle, P. L., J. L. Alderfer, et al. (1997). "Three-dimensional structure of the cytoplasmic face of the G protein receptor rhodopsin." Biochemistry 36(32): 9649-54.

Yeagle, P. L., J. L. Alderfer, et al. (1997). "The first and second cytoplasmic loops of the G-protein receptor, rhodopsin, independently form beta-turns." Biochemistry 36(13): 3864-9.

Yu, H., M. Kono, et al. (1995). "A general method for mapping tertiary contacts between amino acid residues in membrane-embedded proteins." Biochemistry 34(46): 14963-9.

Yu, L., C. Sun, et al. (2005). "Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex." Biochemistry 44(48): 15834-41.

Zhao, J., H. Zheng, et al. (2006). "NMR characterization of recombinant transmembrane protein CB2 fragment CB2(180-233)." Protein. Pept. Lett. 13(4): 335-42.

Zheng, H., J. Zhao, et al. (2006). "A transmembrane helix-bundle from G-protein coupled receptor CB2: biosynthesis, purification, and NMR characterization." Biopolymers 83(1): 46-61.

Zhou, F. X., H. J. Merianos, et al. (2001). "Polar residues drive association of polyleucine transmembrane helices." Proc. Natl. Acad. Sci. U S A 98(5): 2250-5.

2. Probing the formation of stable tertiary structure in a model mini-protein at atomic resolution: Determinants of stability of a helical hairpin

The minimal model system to study the basic principles of protein folding is the hairpin. The formation of β-hairpins, which are the basic components of antiparallel β-sheets, has been studied extensively in the past decade but much less is known about helical hairpins. Here, we probe hairpin formation between a polyproline type-II helix and an α-helix as present in the natural mini-protein peptide YY (PYY). Both turn sequence and interactions of aromatic side-chains from the C-terminal α-helix with the pockets formed by N-terminal Pro residues are shown by site-directed mutagenesis and solution NMR spectroscopy in different solvent systems to be important determinants of backbone dynamics and hairpin stability, suggesting a close analogy with some β-hairpin structures. It is shown that multiple relatively weak contacts between the helices are necessary for the formation of the helical hairpin studied here, whilst the type I β-turn acts like a hinge, which through certain single amino acid substitutions is destabilized such that hairpin formation is completely abolished. Denaturation and renaturation of tertiary structure by temperature or co-solvents were probed by measuring changes of chemical shifts. Folding of PYY is both reversible and cooperative as inferred from the sigmoidal denaturation curves displayed by residues at the interface of the helical hairpin. Such mini-proteins thus feature an important hallmark of globular proteins and should provide a convenient system to study basic aspects of helical hairpin folding that are complementary to those derived from studies of β-hairpins.

published as: Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R., Zerbe, O. (2007). J. Am. Chem. Soc. 129(28): 8811-7

2.1. Introduction

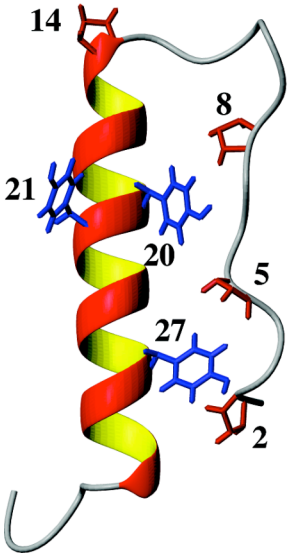
Despite extensive efforts in studying folding of peptides and proteins, a much better understanding of molecular determinants for particular folds and their formation is still highly desirable. While polypeptide chains that fold into specific secondary structures are comparatively easy to design, the construction of proteins with defined tertiary structure, good side-chain packing and cooperative folding behavior is still a major challenge. The reason for this difficulty is that tertiary contacts are often stabilized by a manifold of weaker interactions clearly underlining the necessity for a deeper understanding of how such contacts are made and stabilized.

The smallest natural domains that fold autonomously into protein-like structures have 32-40 residues (Cowley, Hoflack et al. 1992; Sudol 1996), although stable β-hairpin structures with a much smaller number of residues have successfully been isolated from a larger protein (Munoz, Thompson et al. 1997) or even designed de novo (Ramirez-Alvarado, Blanco et al. 1996). The availability of small and well-folded β-hairpin structures is presently advancing our understanding of factors that govern protein folding at ever-increasing pace (see e.g. Hughes et al. (Hughes and Waters 2006) for a review). Particular interest has been paid to β-hairpins that display cooperative folding behavior, since this is a hallmark of natural proteins (Fersht 1999). It has been found that the stability of β-hairpins is determined by contributions from turn (de Alba, Jimenez et al. 1997; Haque and Gellman 1997), intrinsic sheet propensities (Phillips, Piersanti et al. 2005) and hydrophobic side-chain interactions across the strands (Ramirez-Alvarado, Blanco et al. 1996). In fact, the relative contributions of intrinsic conformational bias and interstrand side-chain – side-chain effects are suggested to be of comparable magnitude (Phillips, Piersanti et al. 2005).

While the formation of individual α-helices and β-hairpins has been studied in great detail, relatively little is known about the factors contributing to helical hairpin formation. In the absence of additional covalent constraints (such as disulfide bonds (Kuroda, Nakai et al. 1994; Du and Gai 2006)) helical hairpins have been reported to be marginally stable (Braisted and Wells 1996) or to oligomerize into larger assemblies such as helical bundles (Betz, Bryson et al. 1995). One of the few sequences shown to fold

into a monomeric and stable hairpin in solution is α t α (Fezoui, Weaver et al. 1994). Although the hydrophobic interface appears to contribute significantly to the stability of the two helices in α t α, thermal unfolding was surprisingly uncooperative. This is indicative of a lack of specific tertiary interactions (by either hydrogen-bonding or interdigitating side-chains) and suggests that α t α represents a molten globule intermediate rather than a protein-like structure.

Here we present a detailed investigation on structural determinants of cooperative helical hairpin formation in 36 residue peptides, which are derived from the neuropeptide Y family of peptide hormones. Some of these peptides adopt a well-defined hairpin structure in water, which was first observed by Blundell (Blundell, Pitts et al. 1981) for avian pancreatic peptide (aPP) using X-ray crystallography. This surprisingly stable helical hairpin is commonly referred to as PP-fold and is characterized by a C-terminal α-helix, which is back-folded via a type I β-turn onto an N-terminal polyproline type-II helix. Tertiary contacts between the α-helix and the polyproline helix are shown in Fig.1:




```
      1              11             21             31
pPYY  YPAKP EAPGE DASPE ELSRY YASLR HYLNL VTRQR Y NH2
pNPY  YPSKP DNPGE DAPAE DLARY YSALR HYINL ITRQR Y NH2
bPP   APLEP EYPGD NATPE QMAQY AAELR RYINM LTRPR Y NH2
```

Figure 1: Structure of porcine PYY with side-chains of Pro and Tyr residues depicted in red or blue, respectively. The sequences of porcine PYY and NPY and bovine PP are displayed on the right. Residues that are conserved among different species are underlined, and residues, which are part of the present mutational analysis, are depicted in red.

The solution structure of the peptide YY (PYY) was shown to possess a highly similar helical hairpin (Keire, Kobayashi et al. 2000; Lerch, Mayrhofer et al. 2004) whilst tertiary structure is surprisingly absent in the highly homologous neuropeptide Y (NPY) (Monks, Karagianis et al. 1996). Despite its small size the peptide YY is stably folded in aqueous solutions, as evidenced by non-averaged NMR parameters, a high-resolution NMR structure and uniformly low internal backbone dynamics, as independently verified from NMR relaxation experiments [17]. To probe the energetic significance of the tertiary contacts we replaced residues at the hydrophobic interface of the hairpin-type structure of PYY by Ala and analyzed by high-resolution NMR spectroscopy the conformational and dynamical properties of the resulting peptides. By varying temperature or the methanol content of the aqueous solvent and monitoring chemical shifts we followed the residue-specific formation of tertiary contacts while changing the physical or chemical environment. The methods used by us efficiently deliver structural and dynamical information on changes in the stabilities and folding behavior of the different mutants at atomic resolution and hence provide a much more precise picture as compared to optical methods such as CD or fluorescence spectroscopy, which are usually limited to monitoring changes on molecular scales only.

The results suggest that helical hairpin formation in PP-fold peptides is both reversible and cooperative and that specific N- and C-terminal tertiary hydrophobic contacts between the polyproline and the α-helix provide the driving force for folding. In addition, structural analysis of substitutions in the turn region indicate that the loop behaves like a hinge, which may (or may not) favor, but does not constrain the hairpin structure. To our knowledge, the present data provide the first detailed investigation on cooperative tertiary structure formation in a natural, stably folded, but otherwise unconstrained helical hairpin and show that PP-fold mini-proteins can serve as a convenient system to study aspects of folding that are complementary to those using β-hairpins.

2.2. Results

The role of residues at the PP-fold interface

In our attempt to understand the contributions to stability of the PP-fold, we first chose to examine contacts at the back-fold interface made between Pro and aromatic Tyr residues. Recently, π-Pro interactions were postulated to contribute substantially to the stability of long-range contacts (Gellman and Woolfson 2002; Bhattacharyya and Chakrabarti 2003). Thus we replaced Pro residues 2,5 and 8 as well as Tyr20, 21 and 27 by Ala to produce all single mutations as well as the quadruple Ala2,5,8,14-PYY mutant by site-directed mutagenesis. The destabilization of the helical hairpin in these mutants is probed by measuring changes in internal backbone dynamics by means of the heteronuclear NOE between the amide nitrogen and its directly attached proton. Fig. 2A displays values of the $^{15}N\{^{1}H\}$-NOE for PYY in solution and when bound to DPC micelles. In the latter environment formation of tertiary structure is efficiently suppressed, demonstrating its utility for quantifying the stability of tertiary structure in the system under study (Bader, Bettio et al. 2001; Bader and Zerbe 2005).

The $^{15}N\{^{1}H\}$-NOE data indicate that none of the Pro mutants are stably back-folded (see Fig. 2B). For the quadruple Pro mutant and NPY the $^{3}J_{HN\alpha}$ scalar couplings are > 6 Hz for all residues in the N-terminal segment, which is indicative of conformational averaging. By contrast, in the single Pro mutants the N-terminal segment is not fully flexible (see Supp. Mat.). Interestingly, the rigidity of the C-terminal α-helix is highly correlated to the rigidity of the turn and the adjacent amino acids encompassing residues 6 to 12. Moreover, the lack of long-range contacts destabilizes in particular the last two turns of the helix.

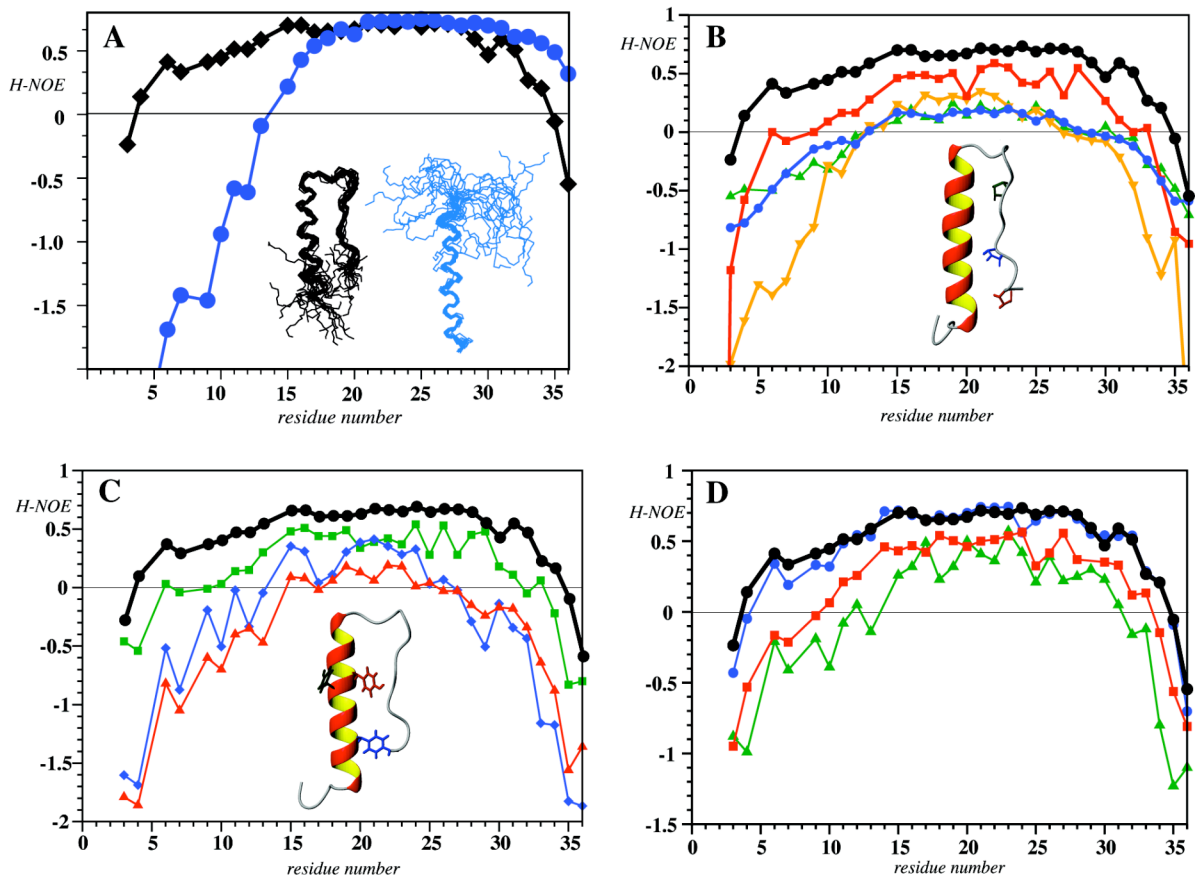


Figure 2: Values of the $^{15}N\{^{1}H\}$-NOE. All data were determined at 500 MHz in water using 1mM solutions of the peptides at pH 4.1 in 20mM acetate buffer. a) Effect of solvent system: Back-folded PYY in aqueous solution (black diamonds) and PYY when bound to DPC micelles (blue circles), where formation of the helical hairpin is suppressed. The corresponding structures displaying backbone atoms are depicted in the inset. b) Effect of proline mutations: PYY (black circles), Ala2-PYY(red squares), Ala5-PYY(blue circles), Ala8-PYY(green triangles) and Ala2,5,8,14-PYY(orange triangles). c) Effect of tyrosine mutations: PYY (black circles), Ala20-PYY (red triangles), Ala21-PYY(green squares) and Ala27-PYY (blue diamonds). d) Effect of mutations in the turn region: PYY(black circles), Ala14-PYY(blue circles), Pro13,Ala14-PYY(red squares) and Ala13-PYY (green triangles).

We have additionally investigated the structural role of Tyr20, 21 and 27 again by recording the $^{15}N\{^{1}H\}$-NOE values of the corresponding single Tyr->Ala mutants (Fig. 2C). The data confirm that these tyrosine residues and Pro2, 5 and 8 have complementary

roles in the stabilization of both secondary and tertiary structure and support the data by Woll and Gellman on the Tyr20Ala and Tyr27Ala mutants (Woll and Gellman 2004). In all the Tyr->Ala mutants, the back-fold is significantly destabilized. Mutation of Tyr20 or Tyr27, however, is far more destabilizing than mutation of Tyr21, because the π-systems of the former two residues point towards the back-folding interface and form contacts with Pro residues 5 and 8. Moreover, we have noticed in case of Ala27-PYY that the heteronuclear NOE drops in particular in the C-terminal part of the destabilized helix encompassing residues 24-30. In contrast, replacing Tyr20, which forms crucial contacts to both Pro5 and Pro8, results in a more general destabilization of the C-terminal helix. Considering that Tyr27 makes contacts with Pro2 removal of the former may result in fraying of the hairpin at its termini.

As the PYY mutants with a significantly destabilized PP-fold lack medium-range proton-proton NOEs in the N-terminal polyproline helix, we chose to probe for residual structure in the N-terminal segment by recording residual dipolar couplings in samples that were partly aligned in stretched polyacrylamide gels. The values measured for the weakly and strongly destabilized PYY mutants Ala2-PYY and Ala8-PYY, respectively, are depicted in Fig. 3 along with the RDCs from the stably back-folded wild-type PYY.

The comparison reveals that overall the RDCs of N-terminal residues of all three molecules follow a similar oscillatory pattern, although the amplitudes are reduced when the helical hairpin is destabilized (as inferred from the $^{15}N\{^{1}H\}$-NOE data). The magnitude of RDC values depends on the ensemble-averaged orientation of the NH bond vectors relative to the alignment tensor. Hence, the decreasing amplitudes indicate that the population of the helical hairpin is reduced in favor of species in which the N terminus is unstructured as seen both in Ala2-PYY and even more pronounced so in Ala8-PYY.

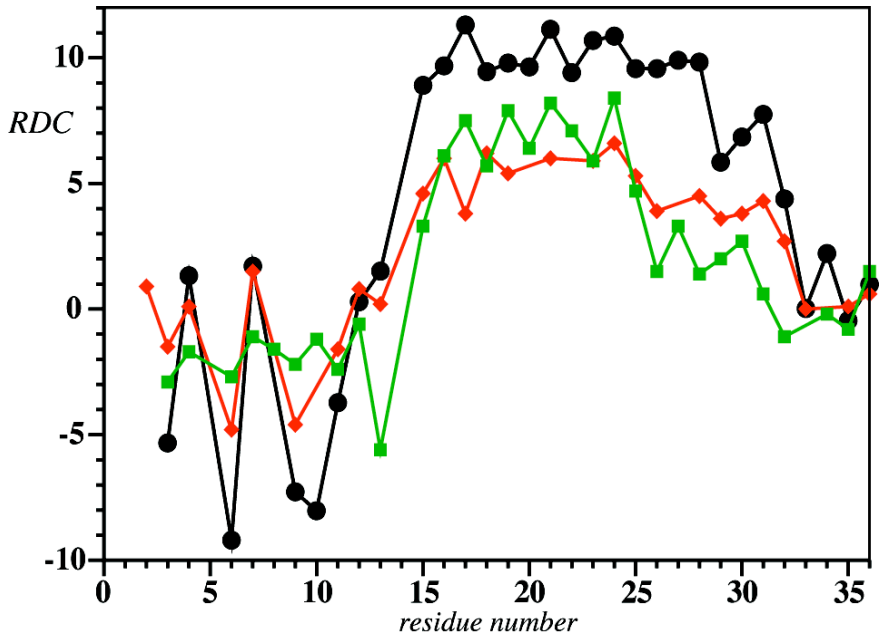


Figure 3: Residual dipolar couplings (RDCs) of PYY(black circles), Ala2-PYY(red diamonds) and Ala8-PYY(green squares) measured in stretched polyacrylamide gels at 700 MHz.

The differences in internal backbone dynamics of the N-terminal segment in different Pro->Ala mutants could possibly be related to i) differences in the propensities of the various mutants to adopt the PP-fold or ii) in changes in the intrinsic rigidity of this peptide segment due to the presence or absence of Pro residues. To resolve this ambiguity we have conducted measurements of the $^{15}N\{^{1}H\}$-NOE in DPC micelles, an environment, in which tertiary interactions are efficiently blocked. We have previously shown in case of bPP or pPYY by using micelle-integrating spin-labels that the amphipathic side of the helix becomes associated with the surface of the micelle, thereby releasing the N terminus (Lerch, Gafner et al. 2002; Lerch, Mayrhofer et al. 2004). In the presence of micelles, the $^{15}N\{^{1}H\}$-NOE data of all single as well as the quadruple mutant are virtually identical (Supp. Mat Fig. S3). Moreover, the C-terminal helix of all peptides is stabilized on binding to micelles, whilst the H-NOE values of residues 6 to 14 indicate continuously increasing flexibility toward the N terminus. A comparison of the H-NOE of the quadruple mutant in water and in DPC micelles displays highly similar values for the N-terminal segment in both environments, providing evidence that the N terminus in

this particular mutant behaves more or less like a free-flight chain also in solution. The data therefore suggest that the backbone rigidity of N-terminal residues of a given mutant in solution is mainly associated with its propensity to adopt the back-fold as a result of a number of specific tertiary contacts between aromatic residues of the C-terminal helix with proline residues of the N-terminal polyproline helix rather than structural properties of the N-terminal segment alone.

The importance of residues in the hinge region

The sequences of pNPY, pPYY and bPP are displayed in Figure 1 along with the structure of pPYY. Pro residues 2, 5 and 8 and Tyr residues 20 and 27, which together form the hydrophobic core of the folded hairpin, are conserved in all three peptides. It is therefore highly unlikely that these residues account for the different propensities of NPY and PYY to adopt the helical hairpin. The sequence homology of NPY and PYY is larger than 80%, and the most prominent difference is Pro found at position 14 in PYY while it occurs at position 13 in NPY. Proline 14 occupies a position in the turn region that links the C-terminal α-helix to the N-terminal polyproline helix. Due to the particular nature of the Pro side-chain we reasoned that Pro14 may direct the N-terminal segment into a favorable enough position for adopting the back-folded state. Accordingly, we have substituted Pro14 by Ala to relax this possible conformational restraint. A comparison of the structure of Ala14-pPYY with pPYY and bPP reveals that Ala14-pPYY very much resembles the structure of PYY or bPP in that the N-terminal segment is clearly back-folded (see Supp. Mat. Fig. S1) and similar interactions between C- and N-terminal residues occur. From these data we deduce that the PP-fold of Ala14-pPYY is sufficiently supported by the hydrophobic contacts made between Pro residues from the N terminus and Tyr residues from the C-terminal α-helix and that Pro14 is not needed to constrain the turn conformationally in order to enforce back-folding.

Although Pro14 apparently does not enforce tertiary contacts in PYY it may be that the otherwise highly homologous NPY cannot adopt the PP-fold because of a Pro in position 13, which is conserved in all NPY sequences known today. The data for the $^{15}N\{^{1}H\}$-NOE of Pro13,Ala14-PYY are depicted in Fig. 2D and clearly prove that this

mutant is not back-folded. A continuous decrease in magnitude of the $^{15}N\{^{1}H\}$-NOE is observed for residues preceding Ala14. Negative H-NOE values are observed for N-terminal residues and the C-terminal helix is also significantly destabilized, but the values also indicate that the hairpin in this mutant is still more stable than in the quadruple mutant. By shifting Pro by one position towards the N terminus the naturally occurring Ser13-Pro14 is replaced by Pro13-Ala14. As a control for the effect of substituting Ser13 alone we also investigated hairpin formation by the single mutant Ala13-PYY. To our surprise we found that the PP-fold is even less stable in Ala13-PYY than in Pro13,Ala14-PYY (see Fig. 2D). An energy minimization of the turn region using an implicit solvent model (see Supp. Mat.) indicates the possible presence of hydrogen bonds of the hydroxyl group of Ser13 with the amide proton of Glu15 and to the side-chain carboxyl group of the latter residue. We conclude that the presence of a hydrogen bond donor at position 13 is crucial for correct formation of the β-turn, and that the shift of Pro from position 14 in PYY to position 13 in NPY or the replacement of Ser13 by Ala is sufficient to disrupt the tertiary contact. It is important to note that, although the amino acid composition of the turn region determines to what extent the turn may favor the formation of tertiary structure, it is nevertheless not sufficient to constrain the helical hairpin on its own. This is evidenced both by the fact that removal of a single tertiary contact in the hydrophobic cluster can fully disrupt the back-fold and by the backbone motional dynamics of the turn region, which is significantly more rigid when the hydrophobic cluster between N- and C-termini is stabilized (see above). On the other hand it is also clear that the turn rigidity weakens gradually under increasingly more destabilizing conditions. This fact is best seen in changes of the $^{3}J_{HN\alpha}$ scalar couplings of residues from the turn region. For example, the coupling for residue Asp11 is always larger than 8 Hz (and often larger than 9 Hz) in the proline mutants in water but always close to 7 Hz in the DPC micelle-bound state, the latter value being indicative of conformational averaging. In addition, the value for this coupling for PYY in methanol is only moderately reduced to 8.7 Hz from the 10 Hz encountered in water. Taken together, these data indicate that the turn acts like a hinge that may support folding but does not provide sufficient stabilizing force to constrain the hairpin on its own.

Studying the formation of the helical hairpin

The stability of the folded state critically depends on the delicate balance between the free energies of the folded and unfolded states. Disrupting the helical hairpin in PYY will expose hydrophobic surfaces to the solvent. Accordingly, we have studied removal of tertiary structure in PYY both by thermal and by solvent-induced denaturation.

In the thermal denaturation experiments, proton spectra of 1mM solutions of PYY and Tyr7-PYY were measured in the temperature range from 280 to 370 K. The corresponding curves are characterized by sigmoidal shapes (see Fig. 4) and the melting points of residues from different regions of the polypeptide chain are within 1 degree for PYY and 1.2 degrees for Tyr7-PYY, suggesting a high degree of cooperativity. The data additionally reveal that the helical hairpin is slightly more stable in Tyr7-PYY than in PYY, as evidenced by an increase of the melting temperature from 316.5K in PYY to 326.5K in Tyr7-PYY:

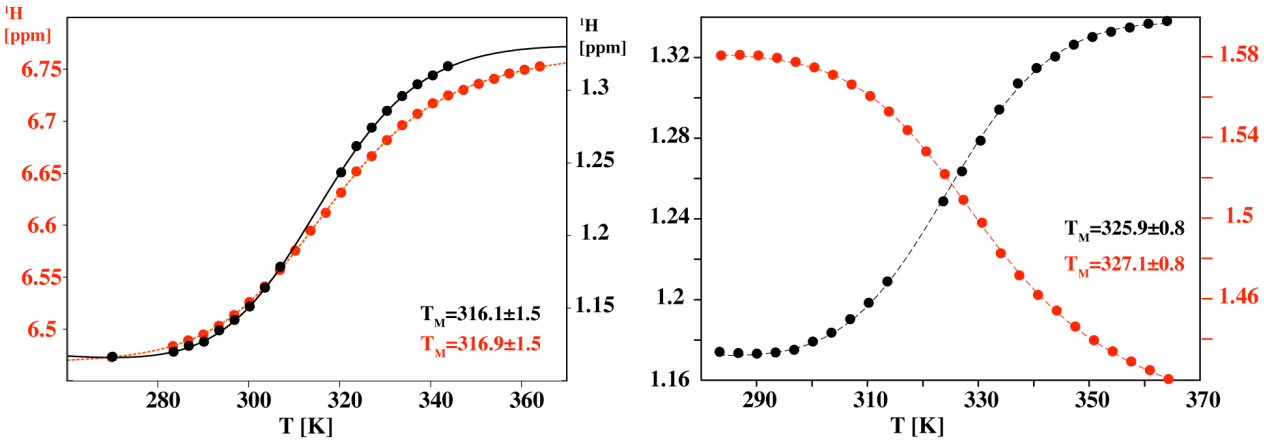


Figure 4: Left: Proton chemical shifts of the methyl group of Ala7 (black circles) and the Hδ of the aromatic system of Tyr20 (red circles) vs. temperature for PYY (left) and of methyl groups of Ala12 (black circles) and Leu17 (red circles) of Tyr7-PYY (right). The melting temperatures derived from fits to experimental data are denoted as T_M in the figure.

The increased PP-fold stability seen in Tyr7-PYY as compared to wildtype PYY must result from additional intra- or intermolecular interactions. It should be noted at this

point that the related avian polypeptide (aPP) exists in dimeric form in the crystal structure (Blundell, Pitts et al. 1981) (K_d 0.32 μM at pH 5.0 (Chang, Noelken et al. 1980)). Therein, the interface between the monomeric subunits is formed by a hydrophobic cluster, in which the π-systems of Tyr7, Tyr21 and Phe20 stack onto one another. By measuring chemical shift changes in a dilution series the K_d for homodimerization of PYY at pH 4.1 was determined to be 35±18 mM, indicating that the monomer concentration is between 91 and 95% at 1mM concentration. The chemical shift of the amide proton of Ala12 is very sensitive to the extent of hairpin formation, and the observed minor change of 0.06 ppm between 1mM and 10μM concentrations indicates that the dimer contributes only to a small extent to the stability of the helical hairpin. We therefore feel justified to treat the system as a monomer.

Apart from thermal denaturation, the PP-fold can also be destabilized by co-solvents with increased lipophilicity. It is known that co-addition of alcohols stabilizes helical secondary structures and disrupts weak hydrophobic contacts, which in turn unfolds tertiary structures. The order of effectiveness for destabilizing tertiary structure is trifluoroethanol > propanol > ethanol > methanol(Bianchi, Rampone et al. 1970; Herskovits, Gadegbeku et al. 1970). The less dramatic changes observed with methanol indicate that this solvent may be particularly useful to monitor the back-folding transition(Kamatari, Konno et al. 1996). Mixtures of water and methanol with methanol contents of about 30% have often been found to result in molten-globule-type structures that possess a considerable extent of native secondary structure while tertiary structure is largely destroyed(de Jongh, Killian et al. 1992; Alonso and Daggett 1995; Bychkova, Dujsekina et al. 1996; Kamatari, Konno et al. 1996; Babu, Moradian et al. 2001; Wang, Ho et al. 2004).

To monitor the back-folding transition we have measured the amide proton chemical shifts as a function of the methanol content. The data for selected residues of Tyr7-PYY, for which the transition between non-back-folded and back-folded species occurs at a larger methanol-to-water ratio than for PYY (and hence is better visible) are displayed in Fig. 5.

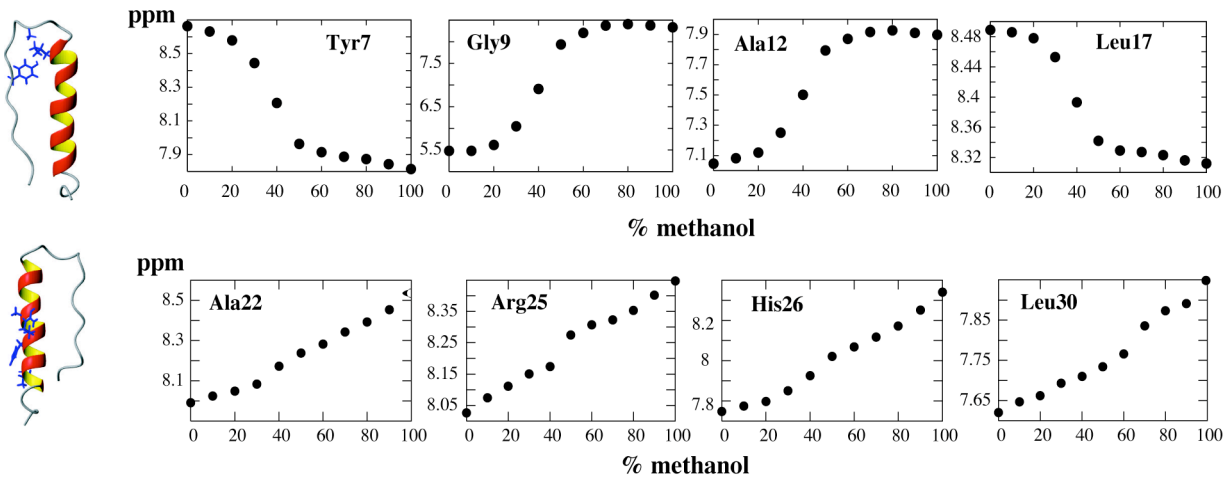


Figure. 5: Proton chemical shifts of selected amide protons of Tyr7-PYY *vs*. the methanol content for residues of the back-fold interface (top) and for those pointing away from it (bottom). On the left, side-chains of these residues have been drawn in the structures to indicate their location in the molecule.

The data clearly reveal that curves from residues of the back-fold interface display a characteristic sigmoidal shape. The point of inflection corresponds to a 41.8 ± 2.1% methanol/water mixture. In contrast, for residues of unstructured segments and for all residues that are pointing away from the back-fold interface the changes are limited to differences in solvent-coordination and/or stability of backbone hydrogen bonds and hence the curves appear approximately linear. The data clearly favor a model in which the PP-fold builds up *cooperatively*, because sigmoidal curves with similar inflection points are observed for the amide proton chemical shift changes for all the residues of the PP-fold interface. Similar sigmoidal curves are also seen for the values of the $^{3}J_{\mathrm{HN}\alpha}$ coupling constants, which are directly dependent on the torsion angle ϕ (see Supp. Mat.). The associated free energy of formation of the helical hairpin in aqueous solution, computed following the procedure introduced by Santoro (Santoro and Bolen 1988), is -17.4 ± 1.3 kJ/mol, which is in agreement with the values typically observed for the free energies of folding of globular proteins ranging from -20 to -60 kJ/mol (Fersht 1999) and similar to the value of -12.6 kJ/mol determined for bPP (Woll and Gellman 2004).

2.3. Discussion

In this study we have investigated molecular properties that lead to the cooperative formation of a helical hairpin in a natural mini-protein at atomic resolution. In particular, we have addressed the importance of the individual Pro and Tyr residues for forming tertiary structure in various mutants of PYY. Moreover, we characterized the significance of residues at specific positions in the turn region. The observation that each Pro->Ala replacement in the segment 2 to 8 reduces the stability of the back-fold significantly indicates that the simultaneous presence of all these contacts is required to form the hairpin. Measurements of backbone dynamics in the presence of DPC micelles, in which the formation of tertiary contacts between the helix and the N-terminal segment is efficiently prevented, revealed that differences in backbone dynamics between the various Pro->Ala mutants in the micelle-bound state are largely removed. Hence the increased rigidity is not primarily an intrinsic property of Pro-rich peptide segments but rather due to correct positioning of residues capable of stabilizing the back-folded conformation.

In a polyproline type II helix positions i and i+3 Pro residues point in the same direction. Moreover, the aromatic residues Tyr20 and Tyr27 are located at i,i+7 positions of a helix and therefore again point in the same direction. Pro residues and aromatic side-chains possess rather flat surfaces, enabling them to mutually intercalate in a zipper-type fashion resulting in a relatively large interaction surface. The importance of non-local Pro-aromatic interactions has been highlightened previously in different contexts, including interstrand stabilization of antiparallel β-sheets [21], a mini-protein called Trp cage (Neidigh, Fesinmeyer et al. 2002; Ding, Buldyrev et al. 2005) and binding of proline-rich peptides to aromatic residues in SH3 domains (Yu, Rosen et al. 1992). It has indeed been speculated that Pro engages with an aromatic residue to form a C-H --- π interaction that may provide substantial binding energy (Gellman and Woolfson 2002; Bhattacharyya and Chakrabarti 2003). It is fully consistent with these observations, that the exchange of individual Tyr residues in the C-terminal α-helix by Ala destabilizes the tertiary structure considerably and that the effects of replacing Tyr20 or Tyr27, which are

forming direct contacts with Pro residues from the N-terminal segment, are much larger than for Tyr21.

Our study additionally revealed that the turn region encompassing residues 10 to 14 is very sensitive to amino acid replacements. Whereas mutating Pro14 in PYY to Ala did not change the rigidity of the helical hairpin significantly, shifting Pro from position 14 to 13 resulted in formation of a non-back-folded species. Therefore it can be safely excluded that Pro14 forces the backbone to adopt a PP-type fold. In contrast, a shift of a single Pro residue by one position in sequence seems to be sufficient to convert peptides that adopt the PP-fold into structures that more closely resemble the conformation of NPY. NPY is clearly not back-folded, although the Pro residues at positions 2, 5 and 8 and Tyr at positions 20 and 27, which together form the hydrophobic core of the hairpin in PYY, are conserved. This points toward a crucial role for the residues of the β-turn region. Interestingly, replacing Ser13 by Ala results in a similar and even stronger destabilization of the helical hairpin. Pro residues are conserved at positions 13 and 14 in NPY and PYY, respectively, and Ser is highly conserved at position 13 in PYY. Computational studies indicate that hydrogen bonds departing from the hydroxyl group of Ser may help to stabilize the β-turn. These observations indicate that residues 13 and 14 serve as a hinge region, and that conformational preferences in that segment are critical to enable stable buildup of the helical hairpin. Although the lifetimes of particular conformational states may be too short to allow characterization through meaningful proton-proton NOEs, our study suggests that backbone dynamics data can be employed for this task. The data reveal that despite significant destabilization of the helical hairpin in some mutants, transient contacts between the N-terminal segment and the helix are formed much more often than it would be expected for a freely diffusing peptide chain. We therefore believe that these mutants sample conformational space similar to what would be expected for a molten globular state, and that they can be considered as models of transient structures being formed during folding of wild-type PYY.

Our data indicate that formation of secondary structure in PYY occurs to some extent independently from buildup of tertiary structure. N-terminal truncation mutants of NPY possess a helix albeit at somewhat reduced stability (Arvidsson, Jarvet et al. 1994). Moreover, NPY possesses no tertiary structure, but is clearly helical, and all the non

back-folded mutants of PYY are still helical. It is therefore reasonable to assume that formation of the C-terminal α-helix in NPY occurs independently and that the tertiary contacts are initiated by hydrophobic contacts made with a pre-formed helix. Once the C-terminal helix is established, however, the PP-fold is formed cooperatively, because contacts of *all* Pro residues in the segment 2 to 8 as well as both Tyr20 and Tyr27 are required. The data suggest that the final conformation is stabilized by a manifold of weak interactions, all of which are required to drive folding to the correct state. This view is strongly supported by the thermal denaturation experiments with PYY and Tyr7-PYY and the measurements of the latter in water-methanol mixtures.

In summary, our results on the folding of an unconstrained helical hairpin suggest that both a turn-promoting sequence and specific non-local Pro-Tyr interactions add to the overall stability of the native fold. While a disfavorable sequence in the β-turn region suffices to destabilize the PP-fold significantly, a favorable sequence alone does not enforce it. Overall, the system should therefore prove useful to investigate the individual roles of the turn and the hydrophobic cluster in the folding dynamics of helical hairpins, for example using ϕ-value analysis (Fersht 1999). Such a study has recently been conducted for a 16-residue β-hairpin from which it was suggested that turn formation is the rate-limiting step during folding, whereas the hydrophobic cluster between the strands slows unfolding. In remarkable analogy to the helical hairpin studied by us it was proposed that main chain and side chain residues of the turn region of this β-hairpin form a characteristic hydrogen bond network, which might play a critical role in stabilizing the folding transition state(Du, Tucker et al. 2006). While we have drawn a comprehensive picture of the molecular interactions underlying the thermodynamics of helical hairpin formation in PYY, studies of the folding kinetics of PYY and the mutants presented here are now needed to shed more light on the mechanistic details of helical hairpin formation.

Conclusions

By NMR spectroscopy and denaturation experiments on a series of mutants we could demonstrate that tertiary structure formation in PYY is strictly dependent on the presence of a number of specific contacts. We could also demonstrate that formation of the PP-fold from preformed secondary structures occurs cooperatively. Most of the Pro->Ala or the Tyr->Ala mutants possess increased backbone dynamics, and the differences in N-terminal mobility among them reflect various degrees to which they sample conformations close to the PP-fold. Molten globules are generally considered being structurally related to the native structure with side-chain conformations less well defined, but many native-like tertiary contacts at least being transiently formed. In that respect many of the mutants studied in this work may actually be considered to be similar to the molten globular states of PYY.

The results may also have implications for our understanding of the binding of these peptides to their cognate membrane–embedded receptors. We have recently postulated that receptor binding is preceded by binding of the ligands to the membrane (Bader and Zerbe 2005). Changes in solvent properties accompany this change in environment. The solvent mixture experiments may therefore simulate the structural transition occurring when a peptide diffuses from the bulk solution towards the membrane. We have noticed that peptide structures are very similar when bound to either micelles or in methanol (Bader and Zerbe 2005) and hence the transition from the membrane-bound state to water may be mimicked by the build-up of tertiary structure in PYY when changing from methanol to water solvent.

2.4. Materials and Methods

All peptides described in this paper have been produced by recombinant methods. They were expressed as insoluble fusions to ketosteroid isomerase, from which they were liberated through cyanobromide cleavage under denaturing conditions. Fusion peptides were expressed in M9 minimal media containing ^{15}N-NH_4Cl as the sole nitrogen source, and verified by their MS and [^{15}N,^{1}H]-HSQC spectra. C-terminal amidation was performed by enzymatic conversion of an extra Gly residue into an amide function using the α-amidating peptidyl glycine amidase (PAM).

The structures of pPYY, Tyr7-PYY and Ala14-PYY in solution were determined following established procedures using 2mM peptide samples at 28°C, pH 4.2 in 20mM deuterated acetate buffer on a Bruker AV-700. 15N relaxation data were recorded at 500 MHz proton frequency on 1mM uniformly 15N labeled samples under similar conditions. Procedures for the spectroscopic work were previously described in more detail (Bader, Bettio et al. 2001; Lerch, Mayrhofer et al. 2004). The structural characterization of the N-terminal polyproline helix proved challenging for two reasons: First, contributions to its stability arise both from medium-range contacts within the segment and long-range contacts to the C-terminal α-helix. Second, whilst a number of long-range proton-proton NOEs between protons of the C-terminal helix and the N-terminal segment clearly define the structure of the N terminus in the back-folded (native) state of PYY, the lack of both medium-range as well as long-range NOEs in the destabilized or denatured states results in great uncertainty under denaturing conditions or in destabilized mutants as to the presence of residual structure of the polyproline helix or transient back-folding to the C-terminal helix. We therefore chose to probe for residual structure by a combination of three different parameters, the ^{15}N{^{1}H}-NOE, $^{3}J_{HN\alpha}$ scalar couplings, and residual dipolar couplings (RDCs) derived from samples that were partially aligned in stretched polyacrylamide gels. We have previously established a relationship between structure and backbone-dynamics that allowed us to quantify the extent of back-folding as an ensemble-averaged quantity of populations of fully back-folded and fully flexible N termini solely based on ^{15}N{^{1}H}-NOE values (Bader, Bettio et al. 2001; Lerch, Mayrhofer et al. 2004). The fact that this model is consistent also with RDCs and

$^3J_{HN\alpha}$ scalar couplings provides further evidence of its validity and usefulness for characterizing folding between two structurally well-characterized states by a single progress variable.

2.5. References

Alonso, D. O. and V. Daggett (1995). "Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water." J. Mol. Biol. 247(3): 501-20.

Arvidsson, K., J. Jarvet, et al. (1994). "Solution structure by 1H and dynamics by natural abundance 13C NMR of a receptor recognising peptide derived from a C-terminal fragment of neuropeptide Y." J. Biomol. NMR 4(5): 653-72.

Babu, K. R., A. Moradian, et al. (2001). "The methanol-induced conformational transitions of beta-lactoglobulin, cytochrome c, and ubiquitin at low pH: a study by electrospray ionization mass spectrometry." J. Am. Soc. Mass. Spectrom. 12(3): 317-28.

Bader, R., A. Bettio, et al. (2001). "Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection." J. Mol. Biol. 305: 307-392.

Bader, R. and O. Zerbe (2005). "Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state?" ChemBioChem 6(9): 1520-34.

Betz, S. F., J. W. Bryson, et al. (1995). "Native-like and structurally characterized designed alpha-helical bundles." Curr. Opin. Struct. Biol. 5(4): 457-63.

Bhattacharyya, R. and P. Chakrabarti (2003). "Stereospecific interactions of proline residues in protein structures and complexes." J. Mol. Biol. 331(4): 925-40.

Bianchi, E., R. Rampone, et al. (1970). "The role of aliphatic alcohols on the stability of collagen and tropocollagen." J. Biol. Chem. 245(13): 3341-5.

Blundell, T. L., J. E. Pitts, et al. (1981). "X-ray analysis (1.4 A resolution) of avian pancreatic polypeptide: small globular protein hormone." Proc. Natl. Acad. Sci. U S A 78: 4175-9.

Braisted, A. C. and J. A. Wells (1996). "Minimizing a binding domain from protein A." Proc. Natl. Acad. Sci. U S A 93(12): 5688-92.

Bychkova, V. E., A. E. Dujsekina, et al. (1996). "Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface." Biochemistry 35(19): 6058-63.

Chang, P. J., M. E. Noelken, et al. (1980). "Reversible dimerization of avian pancreatic polypeptide." Biochemistry 19(9): 1844-9.

Cowley, D. J., J. M. Hoflack, et al. (1992). "Structure of neuropeptide Y dimer in solution." Eur. J. Biochem. 205(3): 1099-106.

de Alba, E., M. A. Jimenez, et al. (1997). "Turn residue sequence determines beta-hairpin conformation in designed peptides." J. Am. Chem. Soc. 119: 175-183.

de Jongh, H. H., J. A. Killian, et al. (1992). "A water-lipid interface induces a highly dynamic folded state in apocytochrome c and cytochrome c, which may represent a common folding intermediate." Biochemistry 31(6): 1636-43.

Ding, F., S. V. Buldyrev, et al. (2005). "Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model." Biophys. J. 88(1): 147-55.

Du, D. and F. Gai (2006). "Understanding the folding mechanism of an alpha-helical hairpin." Biochemistry 45(44): 13131-9.

Du, D., M. J. Tucker, et al. (2006). "Understanding the Mechanism of b-Hairpin Folding via F-Value Analysis." Biochemistry 45: 2668-2678.

Fersht, A. (1999). Structure and mechanism in protein science. New York, W. H. Freedman and Company.

Fezoui, Y., D. L. Weaver, et al. (1994). "De novo design and structural characterization of an alpha-helical hairpin peptide: a model system for the study of protein folding intermediates." Proc. Natl. Acad. Sci. U S A 91(9): 3675-9.

Gellman, S. H. and D. N. Woolfson (2002). "Mini-proteins Trp the light fantastic." Nat. Struct. Biol. 9(6): 408-10.

Haque, T. S. and S. H. Gellman (1997). "Insights on beta-hairpin stability in aqueous solution from peptides with enforced type I' and type II' -turns." J. Am. Chem. Soc. 119: 2303-2304.

Herskovits, T. T., B. Gadegbeku, et al. (1970). "On the structural stability and solvent denaturation of proteins. I. Denaturation by the alcohols and glycols." J. Biol. Chem. 245(10): 2588-98.

Hughes, R. M. and M. L. Waters (2006). "Model systems for beta-hairpins and beta-sheets." Curr. Opin. Struct. Biol. 16(4): 514-24.

Kamatari, Y. O., T. Konno, et al. (1996). "The methanol-induced globular and expanded denatured states of cytochrome c: a study by CD fluorescence, NMR and small-angle X-ray scattering." J. Mol. Biol. 259(3): 512-23.

Keire, D. A., M. Kobayashi, et al. (2000). "Solution structure of monomeric peptide YY supports the functional significance of the PP-fold." Biochemistry 39(32): 9935-42.

Kuroda, Y., T. Nakai, et al. (1994). "Solution structure of a de novo helical protein by 2D-NMR spectroscopy." J. Mol. Biol. 236(3): 862-8.

Lerch, M., V. Gafner, et al. (2002). "Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles." J. Mol. Biol. 322(5): 1117-33.

Lerch, M., M. Mayrhofer, et al. (2004). "Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors." J. Mol. Biol. 339(5): 1153-68.

Monks, S. A., G. Karagianis, et al. (1996). "Solution structure of human neuropeptide Y." J. Biomol. NMR 8(4): 379-90.

Munoz, V., P. A. Thompson, et al. (1997). "Folding dynamics and mechanism of beta-hairpin formation." Nature 390(6656): 196-9.

Neidigh, J. W., R. M. Fesinmeyer, et al. (2002). "Designing a 20-residue protein." Nat. Struct. Biol. 9(6): 425-30.

Phillips, S. T., G. Piersanti, et al. (2005). "Quantifying amino acid conformational preferences and side-chain-side-chain interactions in beta-hairpins." Proc. Natl. Acad. Sci. U S A 102(39): 13737-42.

Ramirez-Alvarado, M., F. J. Blanco, et al. (1996). "De novo design and structural analysis of a model beta-hairpin peptide system." Nat. Struct. Biol. 3(7): 604-12.

Santoro, M. M. and D. W. Bolen (1988). "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants." Biochemistry 27(21): 8063-8.

Sudol, M. (1996). "Structure and function of the WW domain." Prog. Biophys. Mol. Biol. 65(1-2): 113-32.

Wang, Y.-F., M.-Y. Ho, et al. (2004). "Using mass spectrometry to probe subtle differences in conformations of several cytochromes c in aqueous and methanol solutions." J. Mass Spectrom. 39: 1523-1530.

Woll, M. G. and S. H. Gellman (2004). "Backbone thioester exchange: a new approach to evaluating higher order structural stability in polypeptides." J. Am. Chem. Soc. 126(36): 11172-4.

Yu, H., M. K. Rosen, et al. (1992). "Solution structure of the SH3 domain of Src and identification of its ligand-binding site." Science 258(5088): 1665-8

2.6. Supplementary material

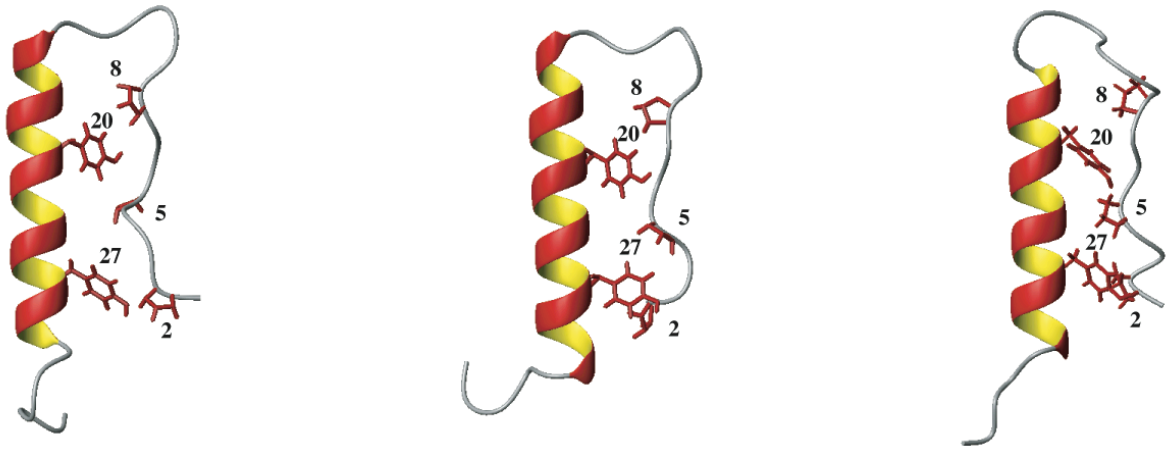


Figure S1: Representative structures of Ala14-pPYY (left), pPYY (middle) and bPP (right).

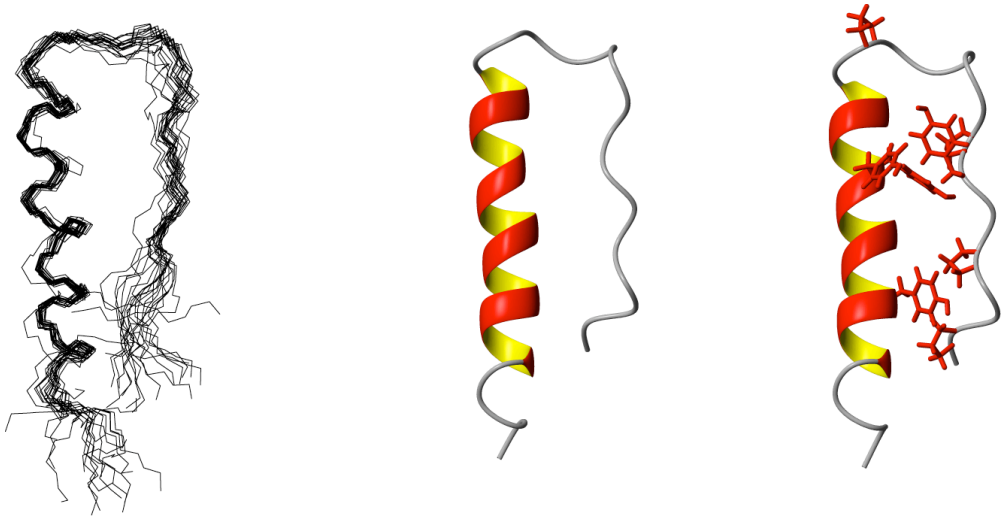


Figure S2: Structure of Tyr7-PYY as determined at pH 4.1, 301K: Backbone presentation of the superposition of single 20 lowest energy conformers (left), and single conformer (middle) with sidechains of Pro and Tyr residues depicted (right).

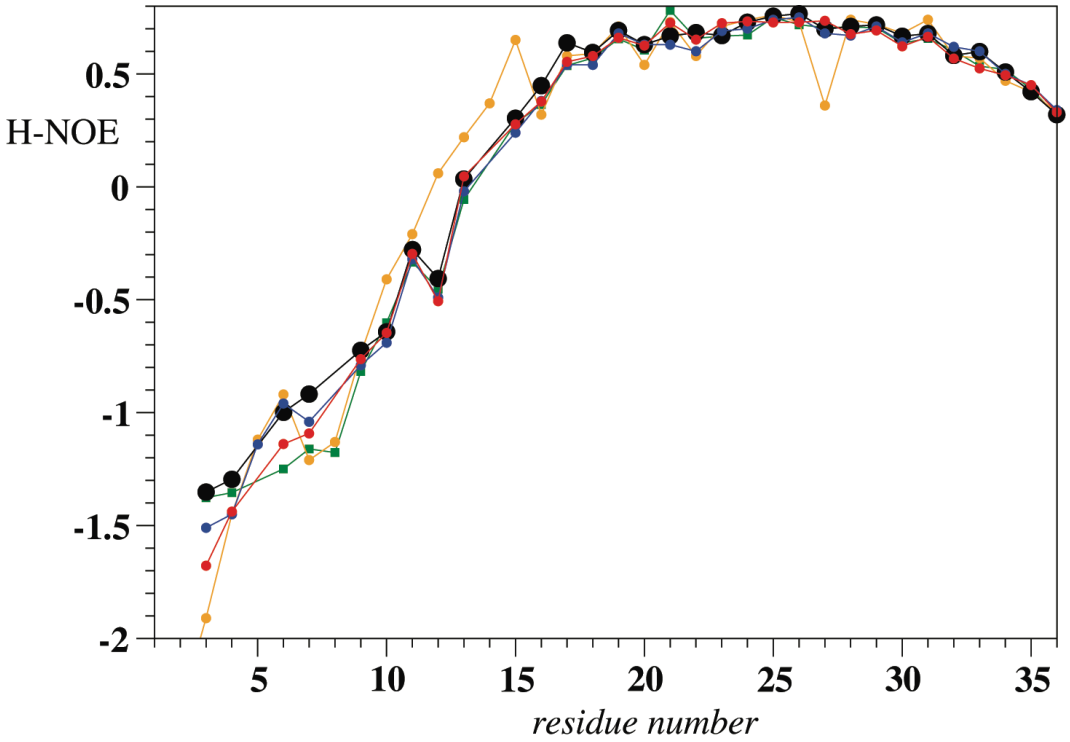


Figure S3: $^{15}N\{^{1}H\}$-NOE of PYY, Ala2-PYY, Ala5-PYY, Ala8-PYY and Ala2,5,8,14-PYY at pH 6.0 in 20 mM MES buffer, 300mM DPC using the same color coding and symbol as in Fig. 2.

Solvent mixture (methanol-water) experiments for Tyr7-PYY:

Mixtures of water and methanol may be particularly useful to study helical hairpin formation in PYY. We have followed the amide proton chemical shifts *vs*. the methanol content. The data for selected residues of Tyr7-PYY, for which the transition between non-back-folded and back-folded species occurs at a larger methanol-to-water ratio compared to PYY and hence is better visible. i) Chemical shifts of residues 3,4,6,7,9,10, 12 17, 18 (23) and 24 of Tyr7-PYY, which are all part of the back-fold interface and hence experience a major structural transition, display a characteristic sigmoidal curve. ii) An approximately linear change occurs for a amide protons being part of an unstructured segment, for which the major change is reflected in the decrease of solvent-coordination upon increasing the content of methanol, e.g. for residues 35 and 36 of Tyr7-PYY as well as from residues 6,7 11, 35 and 36 of PYY. The resonance frequencies of these protons shift to higher field with increasing methanol content. iii) Amide proton resonances of

residues 20, 22, 25, 26, 29 and 30-32, all of which are pointing away from the back-folding interface, are shifted to lower field with increasing methanol content. These changes can be attributed to the fact that the isolated helix and hence the involved hydrogen bonds are stabilized in methanol.

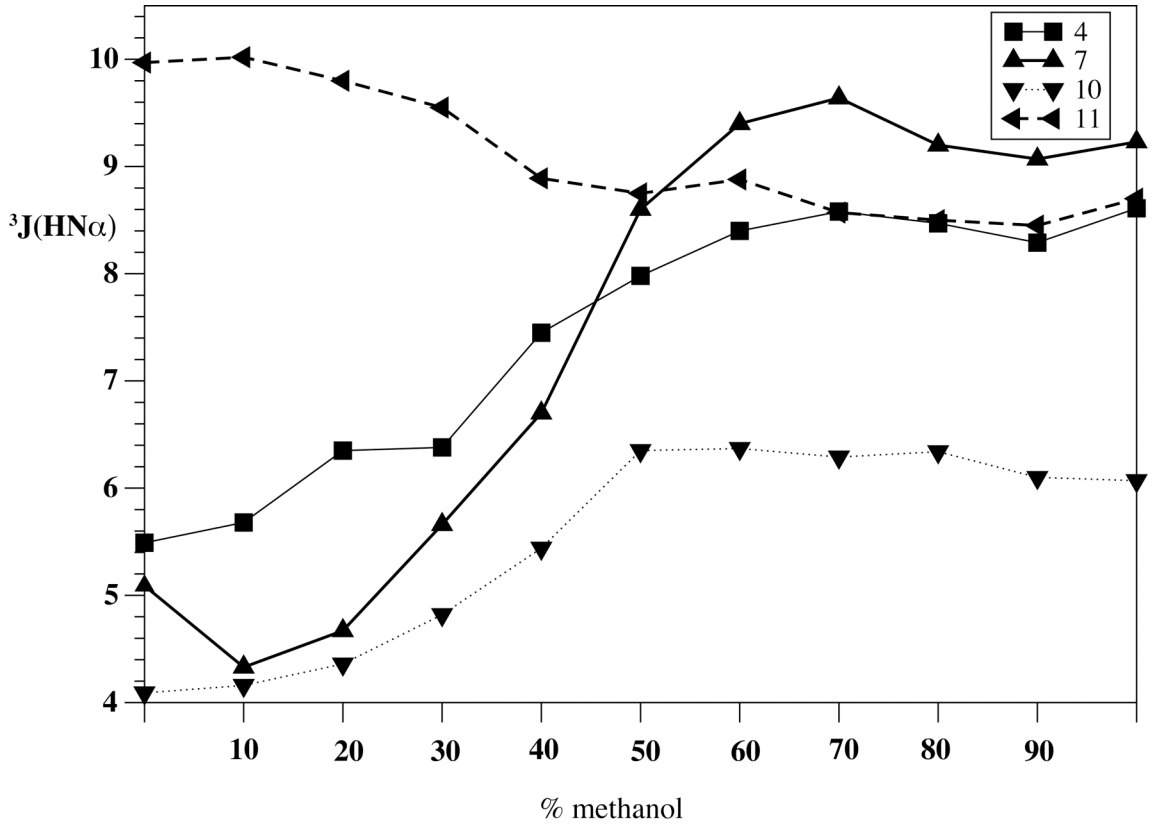


Figure S4: ^{3}J(HNα) scalar coupling constants for Tyr7-PYY in various water-methanol mixtures.

The data display sigmoidal curves. Interestingly, while the H-NOE data indicate that the N-terminal segment is largely flexible at 100% methanol, the scalar coupling data are only compatible with incomplete averaging of dihedral angles. While further work in our laboratory is in progress to investigate this issue in more detail, we speculate that the presence of Pro residues restricts conformational space for backbone rotameric states.

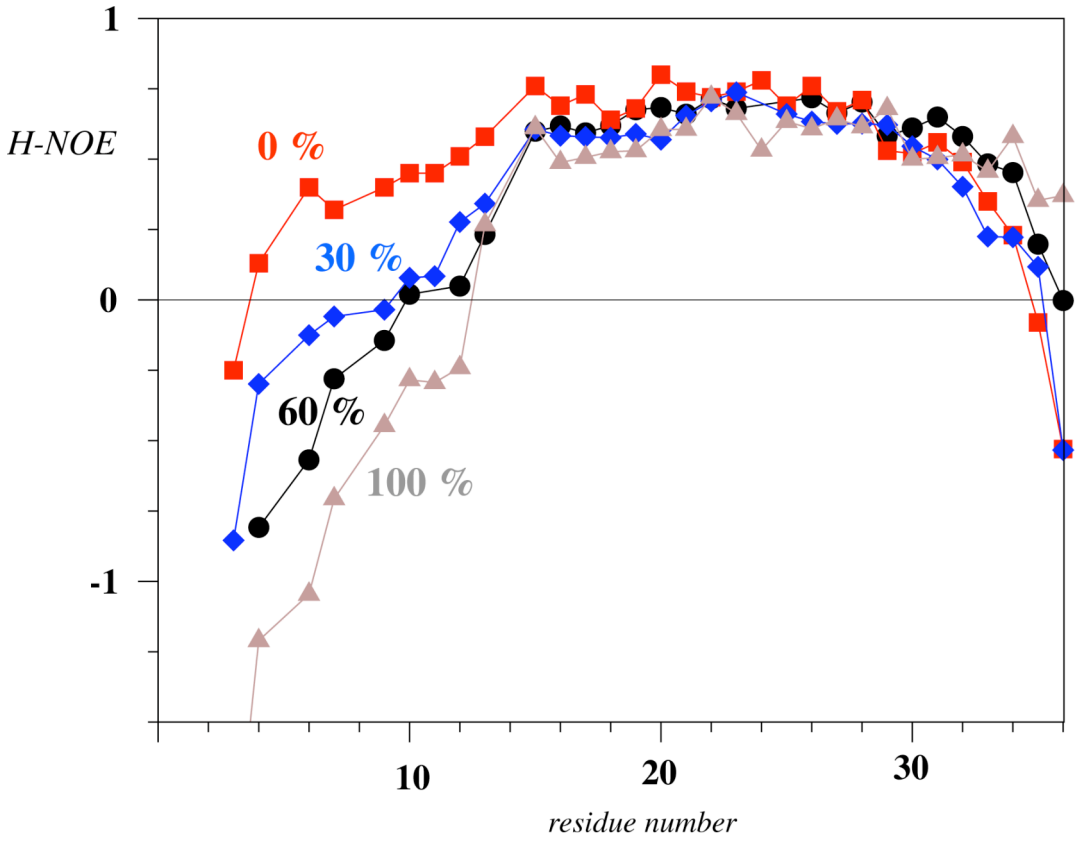


Figure S5: H-NOE data of PYY in various water-methanol mixtures (percentage methanol indicated).

Dimerization of PYY and Tyr7-PYY

It should be noted that the related avian polypeptide (aPP) exists in dimeric form in the crystal structure. Therein, the interface between the monomeric subunits is formed by a hydrophobic cluster, in which the π-systems of Tyr7, Tyr21 and Phe20 stack onto each other. In PYY position 7 is occupied by Ala, and hence we decided to replace it by Tyr. The solution structure of Tyr7-PYY and the $^{15}N\{^{1}H\}$-NOE data indicate that the backbone of Tyr7-PYY is very similar to PYY, and slightly less rigid in the turn region comprising residues 10 to 14. The K_d for homodimerization of PYY is reported to be 22 mM at pH 5. It is difficult to measure K_d's in that range due to the limited solubility of the peptide at very high concentrations, but dilution NMR experiments conducted with PYY suggest a similar value at pH 4.1 (35 ±18 mM). It is also difficult to follow changes of backbone dynamics at very low concentrations to derive estimates of the contribution

of dimerization to the stability of the back-fold. However, it is still possible to monitor chemical shift changes upon dilution. The amide proton frequency of Ala12 is very sensitive to the extent of back-folding: While it is at about 7.3 ppm in PYY, it is measured to be 7.9 ppm in the strongly destabilized mutant Ala8. The amide proton shift of Ala12 shifts by -0.06 ppm upon dilution from 1mM to 20uM. At the latter concentration the population of the monomer is higher than 99%, while it is between 91% and 95% at 1mM for K_d's in the range between 20-40mM. The data indicate that the dimer contributes to the stability of the helical hairpin only to a small extent, so that the gross effects must be due to changes within the monomer. We therefore feel justified to treat the system as a monomer to simplify the analysis.

Turn geometry optimizations and energy calculations

The turn segments of aPP and PYY comprising residues 11 – 15 were geometry optimized both in fully extended and in their native conformations. For the optimization of the native turn conformation, the coordinates of the backbone heavy atoms of residues 11 and 15 were held fixed in the positions as initially derived from the crystal structure of aPP (1) and the NMR structure of PYY (2), respectively. Geometry optimizations were performed using the B3LYP method (3) and the lacvp double-ζ quality basis set, as implemented in the Jaguar 4.2 program (4). Polarization effects from the surrounding solvent were taken into account by using a dielectric cavity model with a dielectric constant of ε=78.8 and a probe radius of 1.4. Table S2 summarizes the electronic energies after optimization of aPP and PYY in extended and turn conformation and the energy difference associated with the conformational change from extended to turn structures.

Table S1

Molecule AA seq 11-15	conformation	el. energy [Hartree]	$\Delta E_{ext->turn}$ [kcal/mole] (error is on the order of 5 kcal/mole)
aPP	extended	-1922,551967	
DAPYE	turn	-1922,561057	-5,70
PYY	extended	-1919,149773	
DASPE	turn	-1919,137608	7,63

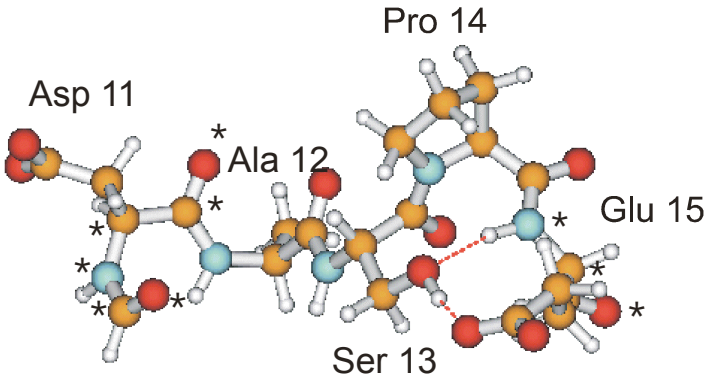


Figure S6: Optimized turn conformation of PYY. Atoms held fixed during the optimization are labeled with a star.

Turn formation in aPP is predicted to be slightly favourable, by contrast to turn formation in PYY, for which calculations predict a slightly unfavourable energy. As mentioned in the manuscript text, turn formation in PYY is not consistent with a Pro in position 13 (by contrast to aPP), nor does it tolerate substitution of Ser13 by Ala. Together, these observations point towards a critical role played by Ser13 in turn formation of PYY, rather than a geometry constraining function exerted by Pro14. One possible mechanism as to how Ser13 may stabilize the turn is evident from Figure S4, showing two hydrogen bonds from and to the side-chain hydroxyl group of Ser13. One is between the hydrogen of the OH group and the side-chain carboxyl group of Glu15 (d=1.56 Å), whilst the other involves the hydroxyl oxygen and the backbone amide group

of Glu15 (d=1.91 Å). Substitution of Ser13 by Ala may therefore remove two turn-supporting interactions and hence critically destabilize the back-folded structure.

References:

(1) Blundell, T. L., Pitts, J. E., Tickle, S. P. & Wu, C. W. (1981). *Proc. Natl. Acad. Sci. U S A* 78, 4175-9.

(2) Lerch, M., Mayrhofer, M. & Zerbe, O. (2004). *J. Mol. Biol.* 339, 1153-68.

(3) Becke, A. D. (1993) *J. Chem. Phys.* 98, 5648-52.

(4) Schrödinger, Inc., Portland, Oregon, Jaguar 4.2, 2000.

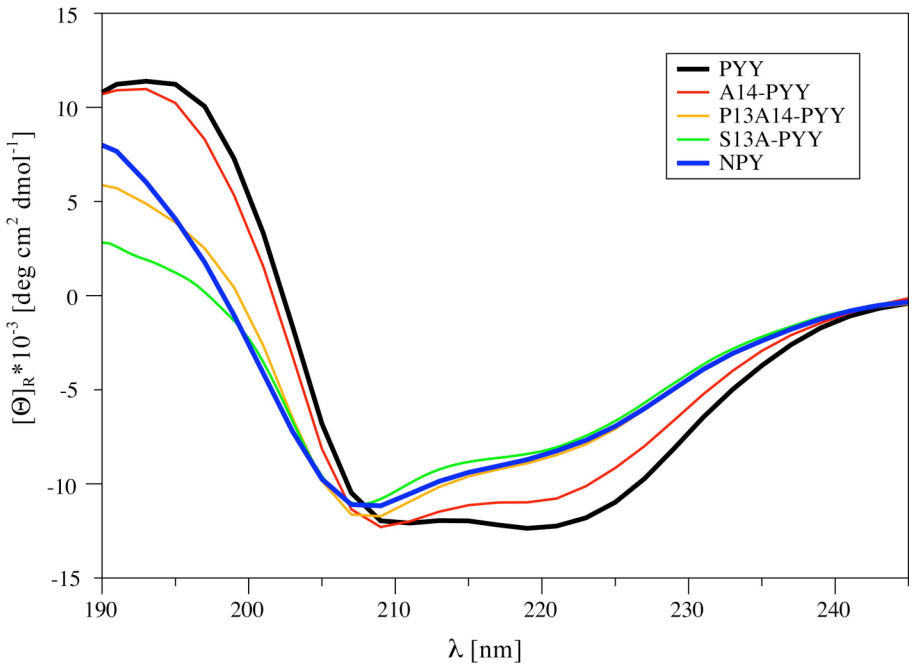


Figure S7: CD spectra of pPYY mutants (α-hairpin turn region), pPYY and NPY. Spectra were recorded in 1mM Na-acetate buffer (pH 4.4) at 298K.

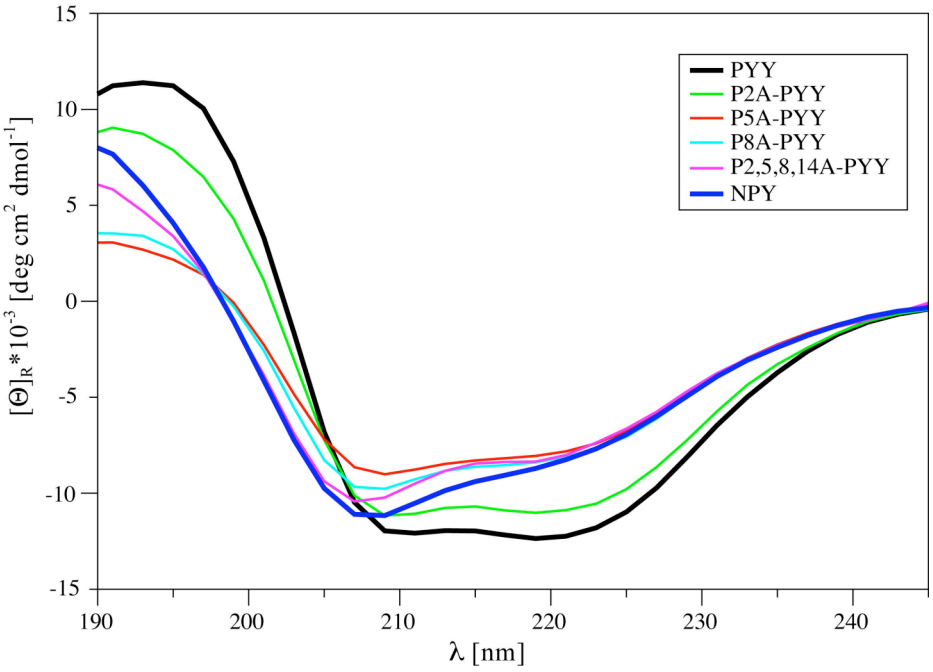


Figure S8: CD spectra of pPYY Pro->Ala mutants (N-terminal segment), pPYY and NPY. Spectra were recorded in 1mM Na-acetate buffer (pH 4.4) at 298K.

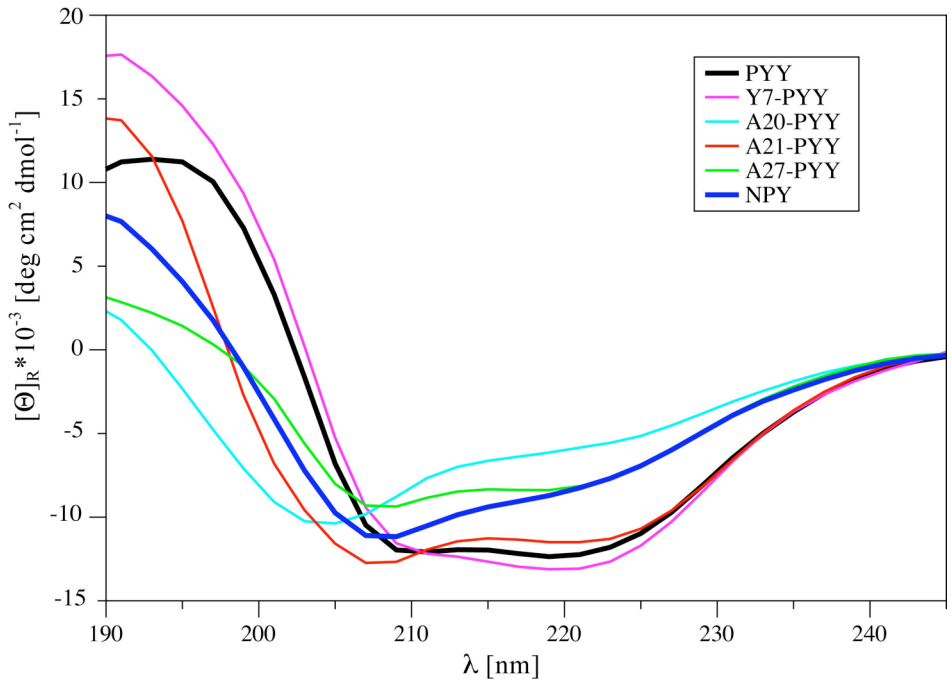


Figure S9: CD spectra of pPYY, Y7-pPYY, pPYY single Tyr->Ala mutants and NPY. Spectra were recorded in 1mM Na-acetate buffer (pH 4.4) at 298K.

Table S2: Chemical shifts for 2mM P14A-pPYY in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm.

N	Res.	H^N	H'^{α}	H^{β}	others
1	TYR	-	4.43	2.97, 3.17	δH 7.18, 7.18; εH 6.84, 6.84; ηOH -
2	PRO		4.47	1.90, 2.29	γCH_2 1.91, 1.91; δCH_2 3.21, 3.65
3	ALA	8.38	4.32	1.33	
4	LYS	8.26	3.68	1.40, 1.57	γCH_2 1.08, 1.08; δCH_2 1.25, 1.25; εCH_2 2.91, 2.91; ζNH_3^+ -
5	PRO		4.34	1.85, 2.27	γCH_2 1.77, 1.77; δCH_2 3.19, 3.50
6	GLU	8.45	4.22	1.85, 1.93	γCH_2 2.36, 2.36; εH -
7	ALA	8.31	3.69	1.11	
8	PRO		4.31	1.77, 2.08	γCH_2 1.61, 1.61; δCH_2 2.93, 2.93-
9	GLY	8.22	3.85, 3.97		
10	GLU	8.34	4.09	1.98, 1.98	γCH_2 2.37, 2.37; εH -
11	ASP	8.51	4.62	2.74, 2.79	δH -
12	ALA	7.46	4.20	1.32	
13	SER	8.35	4.40	4.01, 4.30	γOH -

14	ALA	8.76	4.09	1.46	
15	GLU	8.49	4.11	1.97, 2.08	γCH_2 2.39, 2.39; εH -
16	GLU	7.87	3.95	1.97, 2.23	γCH_2 2.32, 2.39; εH -
17	LEU	8.40	3.97	1.56, 1.82	γH 1.32; δCH_3 0.81, 0.81
18	SER	8.28	4.26	4.02, 4.02	γOH -
19	ARG	7.95	4.11	1.81, 1.91	γCH_2 1.60, 1.60; δCH_2 3.18, 3.18; εNH 7.34; ηNH_2 -, -
20	TYR	8.05	4.33	2.95, 3.12	δH 6.65, 6.65; εH 6.47, 6.47; ηOH -
21	TYR	8.64	3.96	2.94, 3.07	δH 7.05, 7.05; εH 6.80, 6.80; ηOH -
22	ALA	7.98	4.11	1.51	
23	SER	8.17	4.29	3.93, 4.03	γOH -
24	LEU	8.60	4.00	1.60, 1.60	γH 1.17; δCH_3 0.82, 0.82
25	ARG	8.08	3.92	1.86, 1.86	γCH_2 1.55, 1.71; δCH_2 3.14, 3.20; εNH 7.24; ηNH_2 -, -
26	HIS	7.87	4.41	3.29, 3.36	$\delta^1 NH$ -; $\delta^2 H$ 7.27; $\varepsilon^1 H$ 8.62; $\varepsilon^2 NH$ -
27	TYR	8.18	4.09	3.14, 3.14	δH 6.96, 6.96; εH 6.69, 6.69; ηOH -
28	LEU	8.55	3.86	1.77, 1.77	γH 1.43; δCH_3 0.82, 0.82
29	ASN	7.96	4.41	2.78, 2.83	δNH_2 6.86, 7.55
30	LEU	7.66	4.07	1.40, 1.64	γH 1.48; δCH_3 0.73, 0.73
31	VAL	7.84	3.82	1.92	γCH_3 0.62, 0.62
32	THR	7.74	4.23	4.15	γCH_3 1.18; γOH -
33	ARG	7.84	4.20	1.82, 1.82	γCH_2 1.62, 1.62; δCH_2 3.13, 3.13; εNH 7.15; ηNH_2 -, -
34	GLN	8.07	4.16	1.86, 1.95	γCH_2 2.28, 2.28; εNH_2 -, -
35	ARG	8.10	4.15	1.63, 1.63	γCH_2 1.36, 1.44; δCH_2 3.06, 3.06; εNH 7.09; ηNH_2 -, -
36	TYR	8.01	4.52	2.85, 3.07	δH 7.09, 7.09; εH 6.76, 6.76; ηOH -

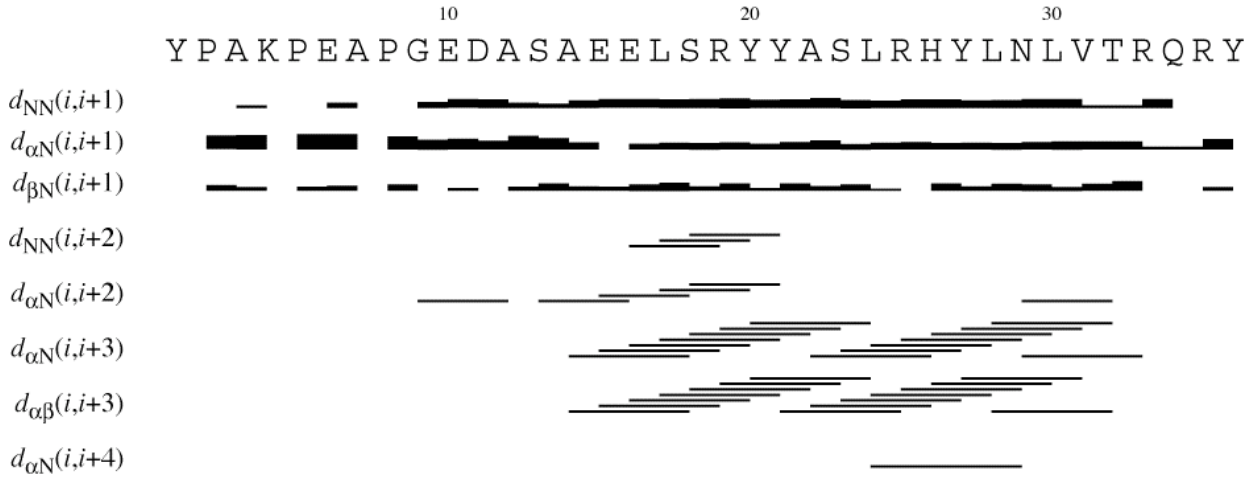


Figure S10: Sequence plot of restraints used during the structure calculation of P14A-pPYY.

Table S3: Information on the structure calculation of P14A-pPYY in solution.

Distance restraints	Total	342
	Intra-residual	84
	Sequential (i - j = 1)	140
	Medium (i - j = 2, 3, 4)	99
	Long-range	19
Dihedral angle restraints		147
RMSD (Å)[a]		
	Pro5-Val31 all heavy atoms	2.48 ± 0.80
	Leu14-Arg29 backbone	0.59 ± 0.30
	Leu14-Arg29 all heavy atoms	1.46 ± 0.29
Structure check[b] (Average %)		
	Tyr1-Tyr36	98.8 (1.2)
NOE constraint violations	Number > 0.1 Å	1
	Maximum (Å)	0.21
Dihedral angle constraint violations	Number > 2.5 degrees	0
AMBER energies (kcal/mol)[c]	Total	-581.8
	Van der Waals	-31.8
	Electrostatic	-875.1

[a] N, C_α, C´ atoms.

[b] Percentage of the ϕ, φ angles falling into the most favored (disallowed) regions within the allowed Ramachandran regions for the 20 refined structures.

[c] AMBER energies are given as the sum of solute-solute and solute-water interactions.

Table S4: Chemical shifts for 1mM A7Y-pPYY in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm.

N	Res.	H^N	$H^{\alpha a}$	H^b	others
1	TYR	-	4.46	3.01, 3.19	δH 7.22, 7.22; εH 6.89, 6.89; ηOH -
2	PRO		4.51	1.98, 2.32	γCH_2 1.91, 1.91; δCH_2 3.20, 3.67
3	ALA	8.38	4.37	1.35	
4	LYS	8.27	3.48	1.35, 1.52	γCH_2 0.94, 1.22; δCH_2 1.59, 1.59; εCH_2 2.93, 2.93; ζNH_3^+ -
5	PRO		4.39	1.88, 2.29	γCH_2 1.70, 1.73; δCH_2 3.08, 3.45
6	GLU	8.54	4.34	1.87, 1.91	γCH_2 2.30, 2.36; εH -
7	TYR	8.6	3.7	2.62, 2.72	δH 6.63, 6.63; εH 6.61, 6.61; ηOH -
8	PRO		4.24	1.70, 2.04	γCH_2 1.26, 1.51; δCH_2 2.37, 3.13
9	GLY	5.45	3.56, 4.25		
10	GLU	8.44	4.17	2.06, 2.06	γCH_2 2.48, 2.48; εH -
11	ASP	8.41	4.73	2.72, 2.87	δH -
12	ALA	7.09	4.16	1.16	
13	SER	8.32	4.68	4.02, 4.39	γOH -
14	PRO		4.28	1.98, 2.41	γCH_2 2.08, 2.22; δCH_2 3.91, 3.96
15	GLU	8.51	4.15	1.97, 2.11	γCH_2 2.40, 2.44; εH -
16	GLU	7.85	4.01	1.90, 2.29	γCH_2 2.36, 2.41; εH -
17	LEU	8.51	4.02	1.55, 1.64	γH 1.87; δCH_3 0.80, 0.89
18	SER	8.48	4.36	4.05, 4.05	γOH -
19	ARG	7.88	4.16	1.95, 1.95	γCH_2 1.64, 1.87; δCH_2 3.18, 3.25; εNH 7.35; ηNH_2 -, -
20	TYR	8.01	4.30	3.08, 3.08	δH 6.65, 6.65; εH 6.56, 6.56; ηOH -
21	TYR	8.73	3.96	2.99, 3.18	δH 7.19, 7.19; εH 6.89, 6.89; ηOH -
22	ALA	7.98	4.16	1.55	-
23	SER	8.24	4.32	3.93, 4.05	γOH -
24	LEU	8.88	3.99	1.03, 1.71	γH 1.54; δCH_3 0.91, 0.98
25	ARG	8.07	3.91	1.75, 1.92	γCH_2 1.59, 1.59; δCH_2 3.18, 3.24; εNH 7.20; ηNH_2 -, -
26	HIS	7.75	4.43	3.34, 3.40	δ^1NH -; δ^2H 7.37; ε^1H 8.65; ε^2NH -
27	TYR	8.24	4.09	3.20, 3.20	δH 6.99, 6.99; εH 6.71, 6.71; ηOH -
28	LEU	8.65	3.87	1.41, 1.77	γH 1.88; δCH_3 0.83, 0.90

29	ASN	7.91	4.44	2.80, 2.87	δNH_2 6.86, 7.52
30	LEU	7.64	4.11	1.41, 1.67	γH 1.52; δCH_3 0.75, 0.75
31	VAL	7.8	3.87	1.95	γCH_3 0.61, 0.64
32	THR	7.72	4.20	4.24	γCH_3 1.22; γOH -
33	ARG	7.89	4.24	1.80, 1.88	γCH_2 1.64, 1.64; δCH_2 3.15, 3.15; εNH 7.14; ηNH_2 -, -
34	GLN	8.10	4.20	1.95, 2.02	γCH_2 2.31, 2.31; εNH_2 -, -
35	ARG	8.12	4.18	1.67, 1.67	γCH_2 1.40, 1.47; δCH_2 3.10, 3.10; εNH 7.09; ηNH_2 -, -
36	TYR	8.04	4.55	2.88, 3.10	δH 7.13, 7.13; εH 6.80, 6.80; ηOH -

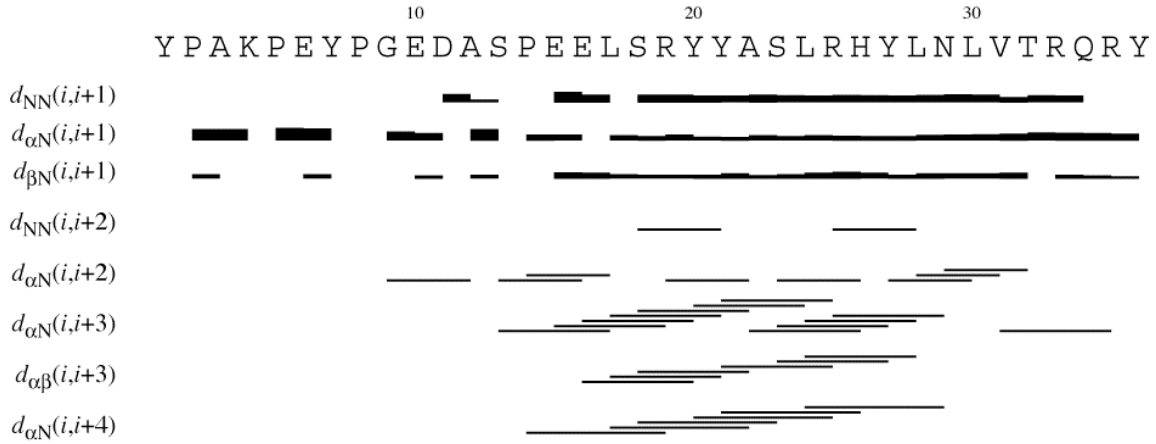


Figure S11: Sequence plot for restraints used during the structure calculation of A7Y-PYY.

Table S5: Information on the structure calculation of A7Y-pPYY in solution.

Distance restraints	Total	552
	Intra-residual	121
	Sequential (i - j = 1)	161
	Medium (i - j = 2, 3, 4)	152
	Long-range	75
Dihedral angle restraints		
RMSD (Å)[a]		
	Pro5-Val31 backbone	$^{1.29}$ ± 0.44
	Pro5-Val31 all heavy atoms	2.21 ± 0.48
	Pro14-Val31 backbone	0.81 ± 0.30
	Pro14-Val31 all heavy atoms	1.72 ± 0.37
Structure check[b] (Average %)		
	Tyr1-Tyr36	99.2 (0.8)
NOE constraint violations	Number > 0.1 Å	0
	Maximum (Å)	-
Dihedral angle constraint violations	Number > 2.5 degrees	0
AMBER energies (kcal/mol)[c]	Total	-1112.0
	Van der Waals	239.1
	Electrostatic	-2561.1

[a] N, C_α, C´ atoms.

[b] Percentage of the ϕ, φ angles falling into the most favored (disallowed) regions within the allowed Ramachandran regions for the 20 refined structures.

[c] AMBER energies are given as the sum of solute-solute and solute-water interactions.

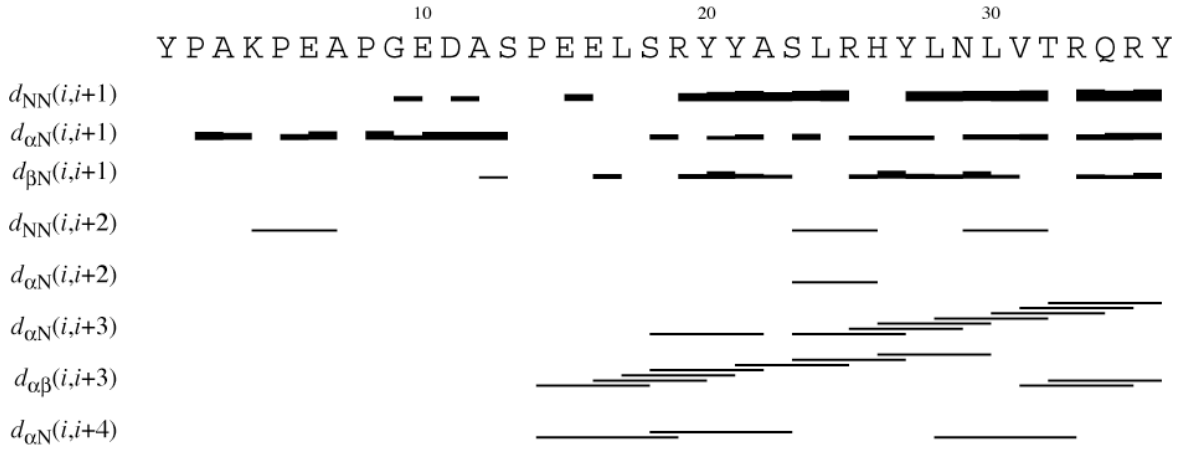


Figure S12: Sequence plot of restraints used during the structure calculation of pPYY in MeOH.

Table S6: Chemical Shifts of pPYY, 1mM, d^3-MeOH, 301K, referenced to residual CD_2**H**OH at 3.31 ppm.

N	Res.	H^N	$H^{\alpha a}$	H^b	others
1	Tyr	-	-	-, -	δH -, -; εH -, -; ηOH -
2	Pro		4.43	1.97, 2.22	γCH_2 -, -; δCH_2 -, -
3	Ala	8.29	4.32	1.38	
4	Lys	8.16	4.63	1.86, 1.86	γCH_2 1.69, 1.69; δCH_2 1.51, 1.51; εCH_2 -, -; ζNH_3^+ -
5	Pro		4.49	-, -	γCH_2 2.09, 2.09; δCH_2 3.65, 3.80
6	Glu	8.29	4.36	1.94, 1.94	γCH_2 2.11, 2.43; εH -
7	Ala	8.07	4.63	1.34	
8	Pro		4.35	1.97, 2.23	γCH_2 -, -; δCH_2 -, -
9	Gly	8.52	3.80, 3.92		
10	Glu	8.07	4.29	2.42, 2.42	γCH_2 2.07, 2.15; εH -
11	Asp	8.27	4.67	2.79, 2.89	δH -
12	Ala	7.95	4.49	1.39	
13	Ser	8.13	4.7	3.98, 4.26	γOH -
14	Pro		4.3	1.94, 2.05	γCH_2 1.69, 1.69; δCH_2 -, -
15	Glu	8.24	4.09	2.21, 2.21	γCH_2 2.51, 2.59; εH -
16	Glu	7.96	4.11	2.17, 2.28	γCH_2 2.51, 2.51; εH -

17	Leu	8.34	4.11	1.86, 1.86	γH 1.50; δCH_3 0.93, 0.93
18	Ser	8.34	4.25	4.01, 4.10	γOH -
19	Arg	7.97	4.06	1.99, 1.99	γCH_2 1.67, 1.67; δCH_2 3.21, 3.21; εNH 7.45; ηNH_2 -, -
20	Tyr	8.44	4.17	3.13, 3.18	δH 6.76, 6.76; εH 6.56, 6.56; ηOH -
21	Tyr	8.67	4.13	3.12, 3.12	δH 7.16, 7.16; εH 6.77, 6.77; ηOH -
22	Ala	8.58	4.05	1.57	
23	Ser	8.23	4.17	4.03, 4.06	γOH -
24	Leu	8.35	4.1	1.68, 1.68	γH 1.50; δCH_3 0.87, 0.87
25	Arg	8.49	3.96	1.99, 1.99	γCH_2 -, -; δCH_2 3.20, 3.20; εNH 7.44; ηNH_2 -, -
26	His	8.4	4.48	3.41, 3.51	δ^1NH -; δ^2H 7.30; ε^1H 8.73; ε^2NH -
27	Tyr	8.49	4.07	3.21, 3.21	δH 6.97, 6.97; εH 6.64, 6.64; ηOH -
28	Leu	8.7	3.92	1.97, 1.97	γH 1.50; δCH_3 0.93, 0.93
29	Asn	8.55	4.38	2.75, 3.10	δNH_2 7.02, 7.79
30	Leu	8.01	3.97	1.74, 1.74	γH 1.60; δCH_3 0.84, 0.84
31	Val	8.27	3.65	1.98	γCH_3 0.75, 0.84
32	Thr	8.03	3.87	4.36	γCH_3 1.28; γOH -
33	Arg	8.03	4.06	1.98, 1.98	γCH_2 1.86, 1.86; δCH_2 3.16, 3.16; εNH 7.44; ηNH_2 -, -
34	Gln	8.15	4.02	2.27, 2.27	γCH_2 2.07, 2.52; εNH_2 -, -
35	Arg	7.89	4.07	1.65, 1.75	γCH_2 1.43, 1.52; δCH_2 3.01, 3.07; εNH 7.25; ηNH_2 -, -
36	Tyr	7.77	4.52	2.84, 3.20	δH 7.22, 7.22; εH 6.67, 6.67; ηOH -

Table S7: Chemical shifts for pPYY mutants, 1 mM samples in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm

N	P2A-pPYY			P5A-pPYY			P8A-pPYY			P2,5,8,14A-pPYY			13P14A-pPYY			S13A-pPYY		
	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm
1	TYR			TYR			TYR			TYR			TYR			TYR		
2	ALA	126.73	8.27	PRO			PRO			ALA	126.96	8.39	PRO			PRO		
3	ALA	124.25	8.25	ALA	124.86	8.37	ALA	124.73	8.37	ALA	124.28	8.25	ALA	124.63	8.37	ALA	124.67	8.37
4	LYS	122.63	8.20	LYS	120.51	8.24	LYS	121.56	8.21	LYS	120.62	8.25	LYS	122.33	8.24	LYS	122.36	8.25
5	PRO			ALA	125.59	8.29	PRO			ALA	125.43	8.30	PRO			PRO		
6	GLU	121.78	8.43	GLU	120.18	8.28	GLU	120.86	8.46	GLU	119.88	8.31	GLU	121.44	8.43	GLU	121.41	8.42
7	ALA	126.91	8.37	ALA	126.03	8.18	ALA	124.91	8.22	ALA	124.98	8.24	ALA	126.85	8.27	ALA	126.81	8.27
8	PRO			PRO			ALA	122.98	8.04	ALA	123.04	8.21	PRO			PRO		
9	GLY	107.85	8.27	GLY	108.88	8.38	GLY	107.78	8.26	GLY	107.78	8.29	GLY	108.21	8.32	GLY	108.28	8.32
10	GLU	118.58	8.33	GLU	119.05	8.24	GLU	119.57	8.18	GLU	119.79	8.17	GLU	118.86	8.23	GLU	118.97	8.22
11	ASP	117.88	8.50	ASP	118.74	8.46	ASP	119.69	8.40	ASP	120.49	8.40	ASP	118.57	8.48	ASP	118.64	8.47
12	ALA	122.74	7.51	ALA	123.31	7.78	ALA	123.54	7.94	ALA	124.55	8.23	ALA	124.18	7.75	ALA	123.08	7.71
13	SER	117.69	8.39	SER	117.05	8.32	SER	116.75	8.25	SER	114.68	8.26	PRO			ALA	124.19	8.27
14	PRO			PRO			PRO			ALA	124.79	8.27	ALA	123.71	8.50	PRO		
15	GLU	117.63	8.48	GLU	117.85	8.45	GLU	118.22	8.44	GLU	118.39	8.25	GLU	117.55	8.67	GLU	118.08	8.73
16	GLU	120.86	7.89	GLU	120.66	7.92	GLU	120.56	7.97	GLU	120.09	8.05	GLU	119.47	7.84	GLU	119.75	7.92
17	LEU	121.03	8.33	LEU	120.79	8.23	LEU	121.09	8.23	LEU	120.97	8.13	LEU	121.57	7.89	LEU	121.65	7.92
18	SER	114.03	8.29	SER	113.99	8.20	SER	114.00	8.13	SER	114.05	8.09	SER	113.88	8.23	SER	114.11	8.24
19	ARG	121.43	7.91	ARG	121.36	7.87	ARG	121.68	7.89	ARG	121.72	7.97	ARG	121.46	7.92	ARG	121.65	7.98
20	TYR	121.71	8.02	TYR	120.85	8.00	TYR	120.98	8.05	TYR	121.01	8.06	TYR	121.22	8.01	TYR	121.06	7.98
21	TYR	118.82	8.53	TYR	118.85	8.36	TYR	119.16	8.31	TYR	119.20	8.31	TYR	119.07	8.36	TYR	119.20	8.31
22	ALA	121.48	8.03	ALA	121.90	8.05	ALA	121.95	8.05	ALA	121.88	8.07	ALA	121.89	8.00	ALA	122.10	8.01
23	SER	115.50	8.10	SER	114.04	7.90	SER	114.24	7.94	SER	114.05	7.93	SER	115.07	8.05	SER	114.93	8.04
24	LEU	120.14	8.38	LEU	122.42	7.64	LEU	122.78	7.77	LEU	122.60	7.70	LEU	123.67	8.14	LEU	123.64	8.11
25	ARG	118.21	7.96	ARG	118.17	7.83	ARG	118.22	7.89	ARG	118.15	7.87	ARG	118.32	7.92	ARG	118.50	7.92
26	HIS	116.69	7.93	HIS	117.12	8.04	HIS	117.23	8.03	HIS	117.25	8.05	HIS	117.03	7.97	HIS	117.16	7.99
27	TYR	120.11	8.14	TYR	119.78	8.05	TYR	119.87	8.07	TYR	119.80	8.07	TYR	120.13	8.13	ALA	120.24	8.13
28	LEU	117.89	8.01	LEU	121.54	8.14	LEU	121.45	8.18	LEU	121.67	8.15	LEU	120.71	8.32	LEU	120.99	8.31
29	ASN	117.98	7.88	ASN	118.39	8.14	ASN	118.39	8.14	ASN	118.43	8.16	ASN	120.85	8.08	ASN	120.84	8.08
30	LEU	120.84	7.77	LEU	121.55	7.92	LEU	121.50	7.91	LEU	121.59	7.94	LEU	118.17	8.06	LEU	118.27	8.08
31	VAL	121.70	8.00	VAL	119.42	7.97	VAL	119.22	7.97	VAL	119.43	7.99	VAL	121.02	7.83	VAL	121.20	7.86
32	THR	115.67	7.87	THR	116.38	7.98	THR	116.29	7.97	THR	116.31	7.99	THR	115.82	7.89	THR	116.09	7.90
33	ARG	121.96	8.02	ARG	122.31	8.12	ARG	122.26	8.10	ARG	122.28	8.12	ARG	122.07	8.04	ARG	122.20	8.06
34	GLN	120.35	8.17	GLN	120.72	8.21	GLN	120.66	8.21	GLN	120.68	8.22	GLN	120.45	8.17	GLN	120.59	8.19
35	ARG	121.23	8.17	ARG	121.60	8.21	ARG	121.55	8.23	ARG	121.57	8.22	ARG	121.35	8.18	ARG	121.48	8.19
36	TYR	120.63	8.06	TYR	120.84	8.08	TYR	120.80	8.07	TYR	120.64	8.03	TYR	120.73	8.06	TYR	120.79	8.06

Table S8: Chemical shifts for pPYY Tyr->Ala mutants, 1 mM samples in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm.

N	Y20A-pPYY			Y21A-pPYY			Y27A-pPYY		
	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm
1	TYR			TYR			TYR		
2	PRO			PRO			PRO		
3	ALA	124.77	8.37	ALA	124.60	8.38	ALA	124.77	8.37
4	LYS	121.49	8.24	LYS	122.91	8.26	LYS	121.34	8.23
5	PRO			PRO			PRO		
6	GLU	120.85	8.42	GLU	122.01	8.46	GLU	120.99	8.41
7	ALA	126.32	8.26	ALA	127.09	8.30	ALA	126.09	8.19
8	PRO			PRO			PRO		
9	GLY	109.08	8.45	GLY	107.92	8.29	GLY	108.78	8.37
10	GLU	119.61	8.13	GLU	118.70	8.31	GLU	119.17	8.22
11	ASP	120.08	8.42	ASP	118.13	8.51	ASP	118.84	8.45
12	ALA	123.86	8.05	ALA	122.75	7.53	ALA	123.35	7.78
13	SER	116.56	8.24	SER	117.45	8.37	SER	117.00	8.30
14	PRO			PRO			PRO		
15	GLU	118.59	8.41	GLU	117.76	8.46	GLU	117.97	8.45
16	GLU	120.40	8.04	GLU	120.96	7.95	GLU	120.67	7.94
17	LEU	121.17	8.13	LEU	120.87	8.35	LEU	121.09	8.24
18	SER	114.52	8.08	SER	116.05	8.31	SER	114.00	8.21
19	ARG	122.10	8.02	ARG	122.72	8.09	ARG	121.47	7.91
20	ALA	122.94	8.05	TYR	120.58	8.15	TYR	121.11	8.01
21	TYR	118.34	8.04	ALA	122.17	8.21	TYR	119.02	8.37
22	ALA	122.63	8.04	ALA	120.12	7.89	ALA	121.91	8.06
23	SER	113.40	8.02	SER	115.96	8.19	SER	114.14	7.90
24	LEU	122.84	7.91	LEU	124.11	8.41	LEU	122.55	7.58
25	ARG	118.91	8.01	ARG	118.51	7.84	ARG	118.36	7.79
26	HIS	117.72	8.14	HIS	116.94	7.99	HIS	117.50	8.09
27	TYR	120.13	8.10	TYR	120.28	8.14	ALA	123.45	8.10

28	LEU	121.94	8.18	LEU	120.17	8.43	LEU	120.41	8.12
29	ASN	118.62	8.19	ASN	118.07	8.02	ASN	118.70	8.25
30	LEU	121.69	7.96	LEU	120.70	7.78	LEU	121.78	8.01
31	VAL	119.54	8.00	VAL	117.52	7.87	VAL	120.31	8.01
32	THR	116.45	8.00	THR	115.54	7.83	THR	117.26	8.08
33	ARG	122.36	8.13	ARG	121.95	8.00	ARG	122.72	8.21
34	GLN	120.76	8.23	GLN	120.31	8.15	GLN	121.11	8.28
35	ARG	121.64	8.22	ARG	121.24	8.16	ARG	121.84	8.25
36	TYR	120.88	8.08	TYR	120.65	8.05	TYR	120.99	8.10

Table S9: Chemical shifts for pPYY Pro->Ala mutants 1 mM samples in 90% H_2O/10%^2H_2O, **300mM DPC**, pH 6.0, T 310K, referenced to the signal of residual HDO at 4.63 ppm.

Res. N	P2A-pPYY			P5A-pPYY			P8A-pPYY			P2,5,8,14A-pPYY		
	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm
1	TYR			TYR			TYR			TYR		
2	ALA	126.88	8.37	PRO			PRO			ALA	126.85	8.39
3	ALA	123.60	8.17	ALA	124.38	8.29	ALA	123.96	8.25	ALA	123.77	8.20
4	LYS	121.01	8.14	LYS	120.07	8.19	LYS	120.56	8.16	LYS	120.15	8.18
5	PRO			ALA	125.35	8.30	PRO			ALA	125.30	8.29
6	GLU	120.55	8.45	GLU	119.76	8.28	GLU	120.56	8.55	GLU	119.70	8.32
7	ALA	125.59	8.18	ALA	125.58	8.19	ALA	124.32	8.19	ALA	124.42	8.18
8	PRO			PRO			ALA	122.88	8.16	ALA	122.92	8.18
9	GLY	108.84	8.40	GLY	108.80	8.39	GLY	107.59	8.27	GLY	107.63	8.27
10	GLU	119.94	8.15	GLU	119.93	8.15	GLU	119.91	8.21	GLU	120.09	8.23
11	ASP	120.53	8.35	ASP	120.49	8.35	ASP	120.41	8.34	ASP	120.78	8.36
12	ALA	123.81	8.12	ALA	123.77	8.11	ALA	123.61	8.09	ALA	124.54	8.27
13	SER	116.38	8.31	SER	116.37	8.31	SER	116.27	8.28	SER	114.72	8.29
14	PRO			PRO			PRO			ALA	124.75	8.28
15	GLU	119.86	8.43	GLU	119.84	8.43	GLU	119.70	8.43	GLU	118.00	8.15
16	GLU	120.69	8.30	GLU	120.68	8.30	GLU	120.63	8.28	GLU	120.26	8.14

17	LEU	119.35	8.31	LEU	119.30	8.30	LEU	119.26	8.30	LEU	118.89	8.28
18	SER	113.19	8.14	SER	113.19	8.14	SER	113.16	8.14	SER	113.15	8.11
19	ARG	122.60	7.92	ARG	122.60	7.92	ARG	122.58	7.91	ARG	122.61	7.92
20	TYR	119.20	8.06	TYR	119.20	8.07	TYR	119.19	8.07	TYR	119.41	8.16
21	TYR	119.13	8.31	TYR	119.12	8.31	TYR	119.18	8.32	TYR	119.44	8.46
22	ALA	120.45	8.19	ALA	120.43	8.19	ALA	120.44	8.19	ALA	120.42	8.20
23	SER	115.20	8.05	SER	115.20	8.05	SER	115.22	8.05	SER	115.28	8.05
24	LEU	122.10	8.24	LEU	122.10	8.24	LEU	122.13	8.25	LEU	122.28	8.30
25	ARG	117.34	8.35	ARG	117.34	8.34	ARG	117.33	8.35	ARG	117.34	8.37
26	HIS	117.43	7.91	HIS	117.44	7.92	HIS	117.32	7.92	HIS	117.37	7.92
27	TYR	117.86	8.13	TYR	117.85	8.13	TYR	117.92	8.14	TYR	118.55	8.16
28	LEU	118.49	8.53	LEU	118.49	8.53	LEU	118.49	8.53	LEU	118.45	8.54
29	ASN	117.71	8.05	ASN	117.72	8.05	ASN	117.66	8.04	ASN	117.68	8.03
30	LEU	118.79	7.66	LEU	118.79	7.66	LEU	118.75	7.66	LEU	118.75	7.66
31	VAL	115.39	7.91	VAL	115.38	7.91	VAL	115.41	7.90	VAL	115.46	7.91
32	THR	110.59	7.83	THR	110.59	7.83	THR	110.60	7.83	THR	110.66	7.83
33	ARG	120.36	7.97	ARG	120.36	7.97	ARG	120.37	7.97	ARG	120.38	7.97
34	GLN	116.62	8.02	GLN	116.62	8.02	GLN	116.62	8.02	GLN	116.63	8.02
35	ARG	118.26	7.89	ARG	118.25	7.89	ARG	118.27	7.89	ARG	118.27	7.89
36	TYR	118.35	7.94	TYR	118.34	7.94	TYR	118.35	7.94	TYR	118.33	7.94

Table S10: $^{15}N\{^{1}H\}$-NOE values for pNPY, pPYY, pPYY mutants and pPYY in MeOH 1mM samples in 90% H_2O/10%$^{2}H_2O$, pH 4.1, T 301K.

Res. N	pNPY	pPYY	P13A14-pPYY	P14A-pPYY	S13A-pPYY	P2A-pPYY	P5A-pPYY	P8A-pPYY	P2,5,8,14A-pPYY	Y20A-pPYY	Y21A-pPYY	Y27A-pPYY	A7Y-pPYY	pPYY in MeOH
2									-2.63					
3	-1.7	-0.24	-0.95	-0.43	-0.88	-1.18	-0.82	-0.55	-1.98	-1.79	-0.46	-1.45	-0.11	-2.04
4	-1.03	0.14	-0.53	-0.05	-0.99	-0.58	-0.78	-0.49	-1.61	-1.86	-0.54	-1.63	0.24	-1.15
5							-0.65		-1.30					
6	-0.8	0.41	-0.16	0.34	-0.21	0.00	-0.49	-0.50	-1.39	-0.82	0.03	-0.48	0.24	-0.91
7	-0.67	0.34	-0.21	0.19	-0.41	-0.07	-0.35	-0.35	-1.27	-1.05	-0.04	-0.85	0.43	-0.69
8								-0.38	-0.95					
9	-0.43	0.41	-0.03	0.33	-0.19	0.00	-0.14	-0.26	-0.81	-0.60	-0.01	-0.16	-	-0.44
10	-0.33	0.45	0.07	0.32	-0.39	0.09	-0.11	-0.32	-0.28	-0.70	0.03	-0.47	0.33	-0.26
11	-0.39	0.51	0.21	0.48	-0.08	0.16	-0.07	-0.20	-0.35	-0.40	0.14	0.10	0.41	-0.25
12	-0.24	0.51	0.25	0.54	0.05	0.17	-0.10	-0.04	-0.11	-0.35	0.15	-0.36	0.53	-0.26
13		0.59		0.53	-0.14	0.28	0.01	0.02	0.06	-0.47	0.30	-0.06	0.51	0.27
14	0.07		0.46	0.71					0.05					
15	0.04	0.70	0.43	0.72	0.26	0.46	0.17	0.10	0.24	0.09	0.48	0.41	0.57	0.60
16	0.15	0.70	0.47	0.69	0.32	0.48	0.16	0.19	0.17	0.08	0.51	0.35	0.66	0.46
17	0.13	0.65	0.42	0.66	0.49	0.49		0.13	0.32	-0.02	0.44	0.10	-	0.49
18	0.17	0.66	0.54	0.68	0.23	0.45	0.12	0.10	0.27	0.06	0.44	0.14	0.60	0.50
19	0.21	0.65	0.50	0.67	0.32	0.50	0.17	0.24	0.31	0.18	0.49	0.34	0.52	0.52
20	0.2	0.67	0.46	0.70	0.50	0.31	0.17	0.14	0.28	0.13	0.34	0.42	0.70	0.57
21	0.39	0.72	0.50	0.74	0.41	0.54	0.19	0.23	0.35	0.06	0.39	0.42	0.70	0.60
22	0.25	0.71	0.51	0.74	0.36	0.59	0.16	0.17	0.31	0.19	0.42	0.40	0.65	0.69
23	0.25	0.69	0.54	0.74	0.57	0.55	0.20	0.22	0.22	0.18	0.37	0.31	0.70	0.63
24	0.23	0.73	0.56	0.56	0.42	0.42	0.16	0.12	0.13	0.01	0.54	0.38	0.66	0.54
25	0.18	0.69	0.32	0.64	0.21	0.41	0.09	0.22	0.19	0.04	0.28	0.04	0.62	0.60
26	0.22	0.71	0.42	0.69	0.39	0.52	0.16	0.15	0.09	-0.03	0.53	0.15	0.64	0.56
27	0.03	0.71	0.56	0.71	0.22	0.31	0.08	0.04	-0.01	-0.03	0.28	0.05	0.64	0.62
28	-0.02	0.69	0.37	0.66	0.25	0.55	0.01	0.00	-0.04	-0.15	0.45	-0.26	0.62	0.58
29	0.1	0.59		0.55	0.30		-0.01	-0.04	-0.07	-0.24	0.48	-0.54	0.64	0.67
30	0.03	0.47	0.35	0.54	0.23	0.27	-0.03	0.05	-0.08	-0.17	0.18	-0.09	0.58	0.52
31	-0.03	0.59	0.33	0.54	0.05	0.10	-0.06	-0.06	-0.21	-0.18	0.11	-0.24	0.45	0.54
32	-0.06	0.51	0.12	0.54	-0.16	0.00	-0.12	-0.05	-0.45	-0.34	-0.05	-0.41	0.30	0.54
33	-0.21	0.27	0.14	0.29	-0.12	0.04	-0.24	-0.28	-0.90	-0.64	0.06	-1.11	0.41	0.44
34	-0.6	0.21	-0.14		-0.80	-0.41	-0.42		-1.22	-0.88	-0.22	-1.31	-0.24	0.56
35	-0.79	-0.06	-0.56	0.21	-1.23	-0.85	-0.59	-0.31	-0.92	-1.56	-0.83	-1.69	-0.68	0.35
36	-0.99	-0.55	-0.81	-0.09	-1.10	-0.95	-0.59	-0.48	-2.63	-1.36	-0.80	-1.60	-0.83	0.42
mean 17-31	0.16±0.09	0.65±0.05	0.46±0.07	0.65±0.06	0.33±0.12	0.43±0.11	0.10±0.07	0.11±0.08	0.14±0.15	0.00±0.14	0.38±0.12	0.11±0.29	0.62±0.07	0.58±0.06

Table S11: $^{15}N\{^{1}H\}$-NOE values for pPYY and pPYY Pro->Ala mutants 1mM samples in 90% H_2O/10%$^{2}H_2O$, **300mM DPC**, pH 6.0, T 310K.

res. N	pPYY	P2A-pPYY	P5A-pPYY	P8A-pPYY	P2,5,8,14A-pPYY
2		-			-2.31
3	-1.35	-1.68	-1.51	-1.38	-1.91
4	-1.30	-1.44	-1.45	-1.35	-1.45
5			-1.14		-1.12
6	-1.00	-1.14	-0.96	-1.25	-0.92
7	-0.92	-1.09	-1.04	-1.16	-1.21
8				-1.18	-1.13
9	-0.72	-0.76	-0.79	-0.82	-0.74
10	-0.64	-0.65	-0.69	-0.60	-0.41
11	-0.28	-0.30	-0.32	-0.33	-0.21
12	-0.41	-0.51	-0.49	-0.45	0.06
13	0.03	0.05	-0.02	-0.06	0.22
14					0.37
15	0.30	0.28	0.24	0.28	0.60
16	0.45	0.38	0.38	0.36	0.32
17	0.64	0.55	0.54	0.54	0.58
18	0.59	0.58	0.54	0.57	0.59
19	0.69	0.66	0.68	0.66	0.71
20	0.63	0.63	0.63	0.61	0.54
21	0.67	0.73	0.63	0.61	0.73
22	0.68	0.65	0.60	0.66	0.58
23	0.67	0.72	0.69	0.67	0.71
24	0.73	0.73	0.70	0.67	0.74
25	0.76	0.73	0.74	0.75	0.76
26	0.77	0.73	0.75	0.72	0.73
27	0.70	0.74	0.68	0.70	0.36
28	0.71	0.68	0.67	0.71	0.74
29	0.72	0.69	0.71	0.70	0.72

30	0.67	0.62	0.64	0.64	0.68
31	0.68	0.66	0.68	0.66	0.74
32	0.58	0.57	0.62	0.60	0.58
33	0.60	0.52	0.60	0.53	0.57
34	0.51	0.49	0.50	0.52	0.47
35	0.42	0.45	0.45	0.43	0.42
36	0.32	0.33	0.34	0.32	0.33
mean 17-31	0.69±0.05	0.67±0.06	0.66±0.06	0.66±0.06	0.66±0.11

Table S12: $^{3}J_{HN\alpha}$ scalar coupling constants for pPYY, pNPY, pPYY mutants in water and pPYY in MeOH. Values ≤ 6.0 Hz are shaded in grey.

Res. N	P2A-pPYY	P5A-pPYY	P8A-pPYY	P2,5,8,14A-pPYY	P14A-pPYY	P13A14-pPYY	S13A-pPYY	pPYY	pNPY	pPYY in MeOH	A7Y-pPYY	Y20A-pPYY	Y21A-pPYY	Y27A-pPYY
1														
2	5.3 ± 0.1			7.5 ± 0.1										
3	6.9 ± 0.1	5.7 ± 0.2	5.7 ± 0.1	6.6 ± 0.1	5.9 ± 0.3	5.7 ± 0.2	5.7 ± 0.2	6.1 ± 0.2	7.3 ± 0.4	5.9 ± 0.1	5.9 ± 0.1	5.3 ± 0.4	5.9 ± 0.4	5.5 ± 0.3
4	6.9 ± 0.1	7.1 ± 0.1	8.8 ± 0.1	7.8 ± 0.1	5.3 ± 0.4	6.2 ± 0.2	6.1 ± 0.2	5.3 ± 0.3	7.7 ± 0.1	7.1 ± 0.1	5.2 ± 0.2	7.4 ± 0.2	6.3 ± 0.2	7.3 ± 0.2
5		6.2 ± 0.2		6.7 ± 0.1										
6	7.6 ± 0.1	7.5 ± 0.1	6.8 ± 0.1	8.1 ± 0.1	6.4 ± 0.4	7.1 ± 0.1	6.9 ± 0.2	7.1 ± 0.2	7.4 ± 0.2	7.7 ± 0.1	7.5 ± 0.2	7.3 ± 0.2	6.9 ± 0.2	7.1 ± 0.3
7	7.1 ± 0.1	5.8 ± 0.2	5.1 ± 0.1	6.7 ± 0.1	5.0 ± 0.4	5.5 ± 0.2	5.4 ± 0.2	5.1 ± 0.3	8.9 ± 0.6	7.6 ± 0.3	4.6 ± 0.3	6.1 ± 0.4	5.1 ± 0.4	6.1 ± 0.4
8			6.1 ± 0.2	6.2 ± 0.1										
9	7.5 ± 0.4	-	6.0 ± 0.3	6.9 ± 0.1	4.3 ± 1.4	6.5 ± 0.3	6.5 ± 0.3	7.4 ± 0.7	6.1 ± 0.7	6.1 ± 0.8	-	5.3 ± 0.9	6.2 ± 0.6	5.5 ± 0.8
10	5.0 ± 0.2	7.6 ± 0.3	6.3 ± 0.1	7.6 ± 0.1	5.1 ± 0.3	5.7 ± 0.2	5.5 ± 0.2	5.5 ± 0.3	7.0 ± 0.2	6.2 ± 0.2	4.7 ± 0.2	6.8 ± 0.3	5.2 ± 0.3	5.6 ± 0.4
11	8.9 ± 0.1	8.1 ± 0.2	8.0 ± 0.1	8.2 ± 0.1	9.5 ± 0.7	9.2 ± 0.1	8.1 ± 0.2	9.8 ± 0.1	9.0 ± 0.1	7.1 ± 0.1	-	7.4 ± 0.3	8.4 ± 0.2	8.4 ± 0.2
12	5.0 ± 0.2	5.6 ± 0.3	6.2 ± 0.2	5.2 ± 0.1	6.1 ± 0.4	5.4 ± 0.2	5.8 ± 0.3	5.0 ± 1.3	6.6 ± 0.6	6.6 ± 0.1	4.7 ± 0.4	6.7 ± 0.3	5.6 ± 0.3	5.5 ± 0.9
13	6.2 ± 0.1	6.1 ± 0.4	6.0 ± 0.1	6.2 ± 0.1	6.0 ± 0.3		5.8 ± 0.2	5.7 ± 0.4		6.4 ± 0.3	5.4 ± 0.2	6.3 ± 0.3	6.0 ± 0.4	6.0 ± 0.4
14				4.6 ± 0.1	4.5 ± 0.7	4.8 ± 0.1			5.0 ± 0.4					
15	4.9 ± 0.1	5.2 ± 0.3	4.6 ± 0.3	5.5 ± 0.1	5.1 ± 0.4	5.6 ± 0.2	5.6 ± 0.3	5.1 ± 0.3	5.2 ± 1.8	5.7 ± 0.3	4.9 ± 0.2	5.4 ± 0.3	5.6 ± 0.3	5.3 ± 0.4
16	5.9 ± 0.2	5.8 ± 0.6	5.8 ± 0.2	5.1 ± 0.1	5.9 ± 0.4	6.5 ± 0.2	6.7 ± 0.3	6.0 ± 0.5	7.1 ± 0.6	6.1 ± 0.2	6.0 ± 0.2	6.5 ± 0.3	5.8 ± 0.7	5.8 ± 0.6
17	5.4 ± 0.2	4.3 ± 0.1	5.2 ± 0.2	5.1 ± 0.1	4.2 ± 0.7	4.7 ± 0.4	5.4 ± 0.4	5.0 ± 0.4	5.7 ± 0.2	5.3 ± 0.2	5.8 ± 0.3	6.2 ± 0.4	5.0 ± 0.7	4.9 ± 0.5
18	4.1 ± 0.2	5.2 ± 0.2	4.9 ± 0.2	4.5 ± 0.1	4.3 ± 0.6	5.2 ± 0.2	4.9 ± 0.3	4.7 ± 0.5	5.2 ± 0.1	5.0 ± 0.2	4.7 ± 0.2	5.9 ± 0.3	5.5 ± 0.3	5.3 ± 0.3
19	6.8 ± 0.3	6.9 ± 0.2	5.1 ± 0.2	5.2 ± 0.1	5.1 ± 0.3	5.7 ± 0.3	5.9 ± 0.4	6.2 ± 0.2	5.2 ± 0.2	5.2 ± 0.2	-	6.0 ± 0.4	4.9 ± 0.9	5.7 ± 0.2
20	-	7.9 ± 0.1	5.4 ± 0.2	5.1 ± 0.1	6.0 ± 0.3	5.2 ± 0.2	5.4 ± 0.2	5.4 ± 0.8	6.2 ± 0.4	5.0 ± 0.2	4.4 ± 0.3	6.0 ± 0.3	4.3 ± 1.3	5.0 ± 0.5
21	4.6 ± 0.4	5.6 ± 0.3	5.5 ± 0.3	5.3 ± 0.2	4.8 ± 0.4	5.1 ± 0.2	5.2 ± 0.3	5.3 ± 1.0	6.6 ± 0.3	4.8 ± 0.2	4.8 ± 0.3	-	5.0 ± 0.4	5.5 ± 0.5
22	5.6 ± 0.4	5.9 ± 0.1	4.8 ± 0.1	4.8 ± 0.1	5.2 ± 0.1	5.0 ± 0.2	5.4 ± 0.2	4.9 ± 0.3	5.8 ± 0.4	4.9 ± 0.2	5.8 ± 0.3	5.8 ± 0.2	5.5 ± 0.3	5.4 ± 0.1
23	5.6 ± 0.1	8.3 ± 0.2	6.0 ± 0.2	5.7 ± 0.1	5.2 ± 0.5	5.7 ± 0.2	5.9 ± 0.4	6.0 ± 1.4	5.8 ± 0.5	5.1 ± 0.2	5.1 ± 0.2	6.1 ± 0.4	4.9 ± 0.9	6.0 ± 0.4
24	5.1 ± 0.1	7.1 ± 0.3	5.3 ± 0.3	6.5 ± 0.4	6.6 ± 0.4	6.0 ± 0.3	6.5 ± 0.3	6.0 ± 0.2	6.3 ± 0.5	5.0 ± 0.2	4.8 ± 0.5	6.3 ± 0.6	6.4 ± 0.4	6.1 ± 0.4
25	6.7 ± 0.1	6.1 ± 0.2	5.7 ± 0.3	6.0 ± 0.2	4.9 ± 0.5	7.1 ± 0.2	6.7 ± 0.3	5.3 ± 0.9	5.4 ± 1.5	5.0 ± 0.2	4.9 ± 0.2	6.6 ± 0.4	5.3 ± 0.4	6.5 ± 0.2
26	5.6 ± 0.3	7.1 ± 0.3	6.3 ± 0.2	6.9 ± 0.1	5.4 ± 0.6	6.0 ± 0.2	6.0 ± 0.4	5.6 ± 0.2	6.5 ± 0.4	4.9 ± 0.2	4.8 ± 0.5	6.6 ± 0.4	5.5 ± 0.4	7.0 ± 0.4
27	7.0 ± 0.1	7.8 ± 0.3	6.0 ± 0.2	6.5 ± 0.1	5.5 ± 0.2	5.8 ± 0.2	6.2 ± 0.3	6.1 ± 0.4	6.5 ± 0.3	4.9 ± 0.2	4.9 ± 0.3	6.3 ± 0.5	-	5.7 ± 0.4
28	6.9 ± 0.5	5.3 ± 0.5	5.7 ± 0.2	6.4 ± 0.1	5.9 ± 0.3	5.6 ± 0.2	5.9 ± 0.3	4.4 ± 1.2	7.1 ± 0.2	4.7 ± 0.2	5.3 ± 0.2	6.7 ± 0.5	5.5 ± 0.5	6.3 ± 0.4
29	5.0 ± 0.2	6.9 ± 0.5	6.8 ± 0.2	7.1 ± 0.1	5.9 ± 0.3	5.0 ± 0.2	-	5.6 ± 0.4	5.9 ± 0.3	5.0 ± 0.2	5.4 ± 0.2	6.8 ± 0.5	5.8 ± 0.8	7.6 ± 0.2
30	6.7 ± 0.1	5.7 ± 0.6	6.8 ± 0.2	6.7 ± 0.2	6.3 ± 0.3	6.3 ± 0.2	6.1 ± 0.3	6.3 ± 0.3	6.5 ± 0.4	5.1 ± 0.2	6.3 ± 0.2	6.9 ± 0.8	6.2 ± 0.5	6.8 ± 0.3
31	7.3 ± 0.1	7.0 ± 0.2	8.7 ± 0.4	7.7 ± 0.1	6.6 ± 0.3	6.4 ± 0.2	6.6 ± 0.3	6.3 ± 0.1	6.6 ± 0.6	5.0 ± 0.2	6.5 ± 0.2	7.3 ± 0.4	6.9 ± 0.3	7.9 ± 0.2
32	6.4 ± 0.1	7.5 ± 0.2	8.6 ± 0.1	7.7 ± 0.1	7.1 ± 0.3	6.3 ± 0.1	7.0 ± 0.3	6.8 ± 0.3	6.9 ± 0.3	5.1 ± 0.2	6.5 ± 0.2	7.2 ± 0.4	6.2 ± 0.3	8.0 ± 0.1
33	6.9 ± 0.1	6.8 ± 0.3	6.2 ± 0.1	7.2 ± 0.1	6.3 ± 0.2	7.2 ± 0.3	6.7 ± 0.3	6.7 ± 0.3	6.9 ± 0.3	5.2 ± 0.2	-	6.6 ± 0.6	6.7 ± 0.4	7.0 ± 0.3
34	-	6.6 ± 0.3	6.2 ± 0.1	7.6 ± 0.1	6.6 ± 0.2	6.4 ± 0.2	6.4 ± 0.2	6.6 ± 0.2	6.0 ± 0.6	5.3 ± 0.2	6.2 ± 0.1	6.7 ± 0.3	7.1 ± 0.6	6.7 ± 0.1
35	7.3 ± 0.1	6.9 ± 0.1	6.7 ± 0.1	7.3 ± 0.1	6.7 ± 0.2	6.9 ± 0.1	6.8 ± 0.2	6.8 ± 0.2	7.5 ± 0.2	6.0 ± 0.2	6.6 ± 0.1	7.4 ± 0.2	6.8 ± 0.3	7.1 ± 0.2
36	8.6 ± 0.1	8.9 ± 0.1	9.0 ± 0.1	8.6 ± 0.1	8.6 ± 0.1	8.8 ± 0.1	8.9 ± 0.1	8.9 ± 0.2	9.3 ± 0.1	9.3 ± 0.1	8.4 ± 0.1	8.6 ± 0.2	8.6 ± 0.2	9.1 ± 0.1

3. Studies of unfolding of PYY in water and methanol by NMR

Unfolding of the 36 amino acid polypeptide PYY was studied by heat- and solvent-induced denaturation. Structural changes were followed using ^{13}C-NMR spectroscopy. Moreover, the structure of PYY in water was re-determined to high resolution using three-dimension ^{13}C-resolved NOESY data as well as a larger set of residual dipolar couplings (RDCs): Moreover, the peptide in methanol solution is characterized both by determining its solution structure as well as by its internal backbone dynamics as derived from ^{15}N relaxation data. The latter is characterized by a complete loss of tertiary structure. The C-alpha chemical shifts in the heat-denaturation experiments display sigmoidal curves with very similar points of inflection indicating that both secondary as well as tertiary structure in the heat denaturation is lost synchronously. In addition, chemical shifts of selected sidechain resonances follow very similar curves, and the disruption of the structure of PYY is triggered by both larger changes in backbone as well as sidechain dihedrals. The data clearly reveal cooperative folding behavior of PYY at single-residue resolution and classify the peptides as a two state-folder. In contrast, the curves during solvent-induced denaturation display a more complicated behavior, which most likely reflects two trends that partially compensate each other: At low low-to-medium contents of methanol tertiary structure is disrupted leading to a loss of stability of secondary structure as well. At high content of methanol secondary structure is again rigidified, as clearly seen in the structure of PYY in pure methanol, which is supported by data on internal backbone dynamics in that solvent.

manuscript in preparation: Mares, J., Walser, R., Neumoin, A., Zou, C., Zerbe, O.

3.1. Introduction

Folding of peptides and proteins has been an intense matter of research in the last decade (Fersht and Daggett 2002). Prediction of the structure from the amino acid sequence is still very difficult, and in the absence of high homology prone to large errors. Therefore, a much better understanding of molecular determinants of particular folds would be highly desirable. Recently, we have proposed that the peptide YY (PYY) presents a convenient system to study basic aspects of protein folding, and for formation of helical hairpins in particular (Neumoin, Mares et al. 2007). The particular fold of PYY was observed by Blundell for avian PP for the first time by crystallography (Blundell, Pitts et al. 1981). It is characterized by a C-terminal α-helix, which is connected via a turn to a N-terminal polyproline type-II helix. A hydrophobic contact between the α-helix and the polyproline helix leads to formation of the hairpin-type structure. Using an extensive set of single points mutants we could recently map the determinants of the PP fold. In particular, we have investigated the role of individual Pro and Tyr residues that are involved in forming the hairpin interface. Our data revealed that the presence of each of these Tyr or Pro residues is crucial, and that replacement of any of them resulted in significant destabilization of the helical hairpin. In addition, mutation of Ser13 in the turn region resulted in complete loss of tertiary structure, additionally emphasizing the role of the amino acid sequence in the turn region.

Recently, much interest has been focused on systems that do not fold cooperatively, and for which different curves are observed for the chemical shifts vs. temperature, e.g. for proteins with small or no barriers for folding that follow so-called downhill folding pathways. The group of Munoz has followed backbone and side-chain proton chemical shifts for the 40-residue polypeptide BBL, which revealed different transition temperatures at different positions along the chain, allowing them to conclude that it belongs to the class of downhill folding proteins (Sadqi, Fushman et al. 2006). To verify that the similarly sized PYY in contrast is really folding in cooperative (two-state) manner we have sought to record a more extensive set of probes. In this work we specifically address this question by investigating the temperature and solvent-induced denaturation of PYY in detail. Accordingly we have expressed PYY in ^{13}C-labelled form

and looked at the temperature profiles of ^{13}C chemical shifts for each residue. We primarily used changes in ^{13}C-alpha chemical shifts that are almost exclusively dependent on ϕ/ψ dihedral angles to characterized changes in backbone conformational preferences.

During the mutagenesis work we have also detected a non-anticipated effect of replacing Ser-13, a residue from the turn region, by Ala (Neumoin, Mares et al. 2007). Unfortunately, the structural reason for this effect was unclear, mainly because the solution structure of PYY was ill-defined in the turn region due to a restraint sampling problem. To improve the precision of the structure we refined it using ^{13}C-resolved NOESY data in combination with a set of residual dipolar couplings. Moreover, the structure of PYY in methanol is also determined, and its structural and dynamical properties compared to the peptide in water. These two structures define the initial and final states of the solvent-induced denaturation process. Changes of structure-dependent parameter such as chemical shifts in combination with these structures allowed deriving a comprehensive picture of unfolding. In addition, it firmly pinpoints cooperativity in folding of PYY in water, and suggests that tertiary structure of PYY can be selectively removed in methanol while only slightly altering secondary structure. Interestingly, for many residues the ^{13}C chemical shift profiles for the solvent denaturation curves are not monomodal, implying more complicated trajectories during the unfolding process. We believe that the data nicely demonstrate cooperativity of folding in water at a resolution that is not accessible using other spectroscopic techniques.

3.2. Results

A refined structure of PYY in water from 13C NOESY data and RDCs

Structures of porcine PYY (pPYY) in solution have been determined by Keire et al (Keire, Kobayashi et al. 2000) and by Lerch et al (Lerch, Mayrhofer et al. 2004). Both structures have confirmed that pPYY assumes a helical hairpin. However, the structure of Keire possessed an unusual kink in the helix around residues, which was absent in the structure from our laboratory. In addition, Poulsen et al. have solved the solution structure of human PYY (Nygaard, Nielbo et al. 2006), and found it to be highly similar to the structure determined by us previously. While the helix in our previous structure was well defined, the fold of the polypeptide sequence in the turn region was much less clear and characterized by higher values of the RMSD. Nevertheless, values of the $^{15}N\{^{1}H\}$-NOE indicated that the turn region is comparably rigid, and indicated that the low resolution is primarily a sampling problem related to the low number of medium-range NOEs in that region. In this work we have now recalculated the solution structure of PYY, using an extended set of upper-distance restraints derived from the additional usage of 3D ^{13}C-NOESY data. The 3D ^{13}C-resolved NOESY in addition to the 2D NOESY resulted in a total of 366 meaningful upper distance constraints compared to 261 in the previously reported structure calculation. Furthermore, additional restraints were derived from measuring residual dipolar couplings in stretched polyacrylamide gels involving 24 $^{15}N,^{1}H$ RDCs, 24 C'-CO RDCs and 10 Cα-Hα RDC (peak overlap and problems with the spectral quality in the Cα-Hα region limited the number of useful correlations).

A comparison of the refined structure of PYY in three different environments, in aqueous solution, in methanol and when bound to dodecylphosphocholine micelles, is depicted in Fig.1. While the RMSDs for backbone and all heavy atoms in the helix region encompassing residues 14-31 after superimposing backbone atoms of the same residues were 0.41±0.15Å (0.78±0.30Å) and 1.53±0.26 Å (1.69±0.39Å), respectively, and hence were slightly but not dramatically improved, the precision in the turn region (residues 5-

14) after fitting backbone atoms of residues 5-31 was significantly increased in the refined dataset with values of 0.93±0.56 Å (vs. 1.77±0.64 Å) and 1.97±0.81 Å (vs. 2.74±0.94 Å) (in brackets the RMSD values in the older data set 1RU5 are shown for comparison).

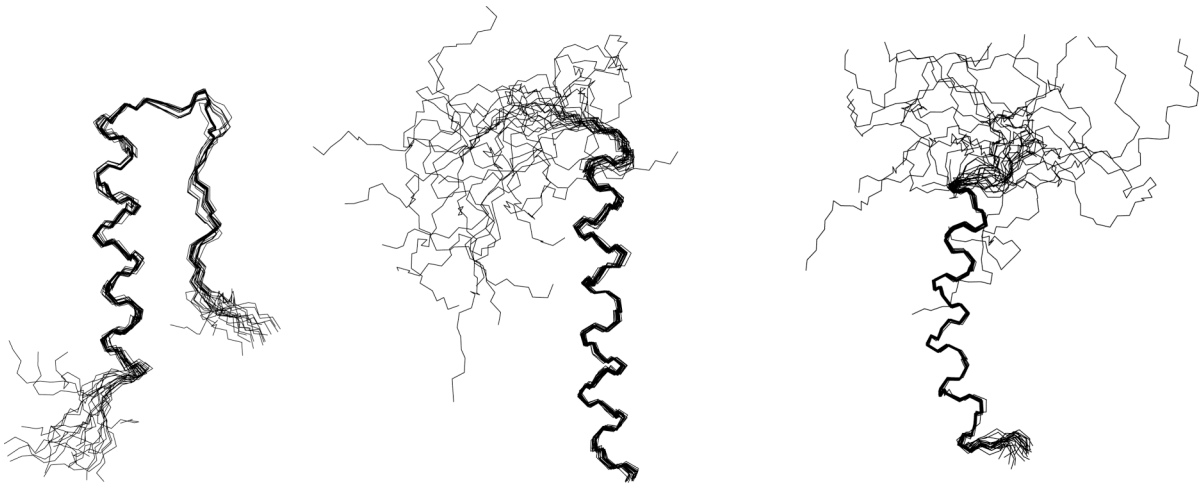


Figure 1: Backbone representation of the superposition of 20 low-energy conformers of PYY in water (left), methanol (middle) and when bound to DPC micelles (right). Backbone atoms of residues 15 to 31 have been superimposed for best fit, for PYY in water residues 4 to 8 have been included in addition.

The much-improved resolution in the β-turn region now allows comparing the solution structure of PYY with the crystal structure of aPP as determined by Blundell in 1981 (Blundell, Pitts et al. 1981). He and his colleagues have postulated the presence of a β I turn in the segment Gly9 to Ala12. Unfortunately, even in the new dataset the dihedrals in that region display significant spread in their values, but average values for phi (and psi) dihedrals of residues 10 and 11 are -62° (-4°) and -85° (23.4°), respectively, and are closer to the ideal values of -60°(-30°) and -90°(0°) of a β 1 turn than to any other β turn. We like to emphasize here that the $^{15}N\{^{1}H\}$-NOE is considerably lower in the turn region (about 0.5) than in the helical region (about 0.7) so that increased flexibility may to some extent account for the larger spread in computed dihedrals.

Recently we have seen that replacement of Ser13 by Ala resulted in complete removal of the helical hairpin[2]. *Ab-initio* calculations indicated that the hydroxyl oxygen from the sidechain might be involved in a hydrogen bond formed with backbone and

sidechain atoms of Glu13. While a few conformers of the energy minimized NMR ensemble did display a H-bond with the backbone amide of Glu16 no such contacts with the sidechain of Glu16 are seen.

The conformation of PYY in methanol is characterized by complete loss of tertiary structure

Unfolding the hairpin in PYY will expose hydrophobic surfaces to solvent, and the free energy of the unfolded state should hence critically depend on the hydrophobicity of the solvent. We have recently reasoned that tertiary contacts in PYY should be destabilized in the presence of a more hydrophobic solvent. We have therefore investigated whether the non back-folded state of PYY could be stabilized when the peptide is dissolved in a solvent of increased lipophilicity such as methanol. Alcohols have been utilized as a solvent for structural studies of amphipatic peptides such as neuropeptide (Mierke, Durr et al. 1992). It has also been realized that co-addition of alcohols stabilizes helical secondary structures while disrupting weak hydrophobic contacts thereby destroying tertiary structures. The order of effectiveness for destabilizing tertiary structure is trifluoroethanol > propanol > ethanol > methanol (Bianchi, Rampone et al. 1970; Herskovits, Gadegbeku et al. 1970). The less dramatic changes observed with methanol indicate that this solvent may be particular useful to monitor the transition (Kamatari, Konno et al. 1996). Mixtures of water and methanol with methanol contents of about 30% have often been found to result in molten-globule-type structures that possess a considerable extent of native secondary structure, but in which the tertiary structure has been largely destroyed (Nakano and Fink 1990; Harding, Williams et al. 1991; de Jongh, Killian et al. 1992; Shiraki, Nishikawa et al. 1995; Bychkova, Dujsekina et al. 1996; Kamatari, Konno et al. 1996; Brutscher, Bruschweiler et al. 1997; Babu, Moradian et al. 2001; Wang, Ho et al. 2004). The group of Daggett has used MD calculation in 60% methanol to simulate unfolding of ubiquitin under these conditions (Alonso and Daggett 1995).

The structure of pPYY in 100% d_3-methanol as calculated from upper-distance limits derived from NOEs using standard 2D NMR methods is depicted in Fig 1, and

compared to the structure of PYY in water or PYY when bound to phospholipids micelles. Obviously, the helical hairpin is completely removed. Not a single NOE supporting an association of the N terminus with the C-terminal helix was observed in the spectra. Nevertheless, the α-helix is stably formed and extends from residues 14 to 33, using the criteria of Kabsch and Sanders (Kabsch and Sander 1983) for identification of secondary structure, and residues 33-36 also occupy helical backbone dihedrals. This is in contrast to the peptide in aqueous solution, where residues 32 to 36 are much more flexible.

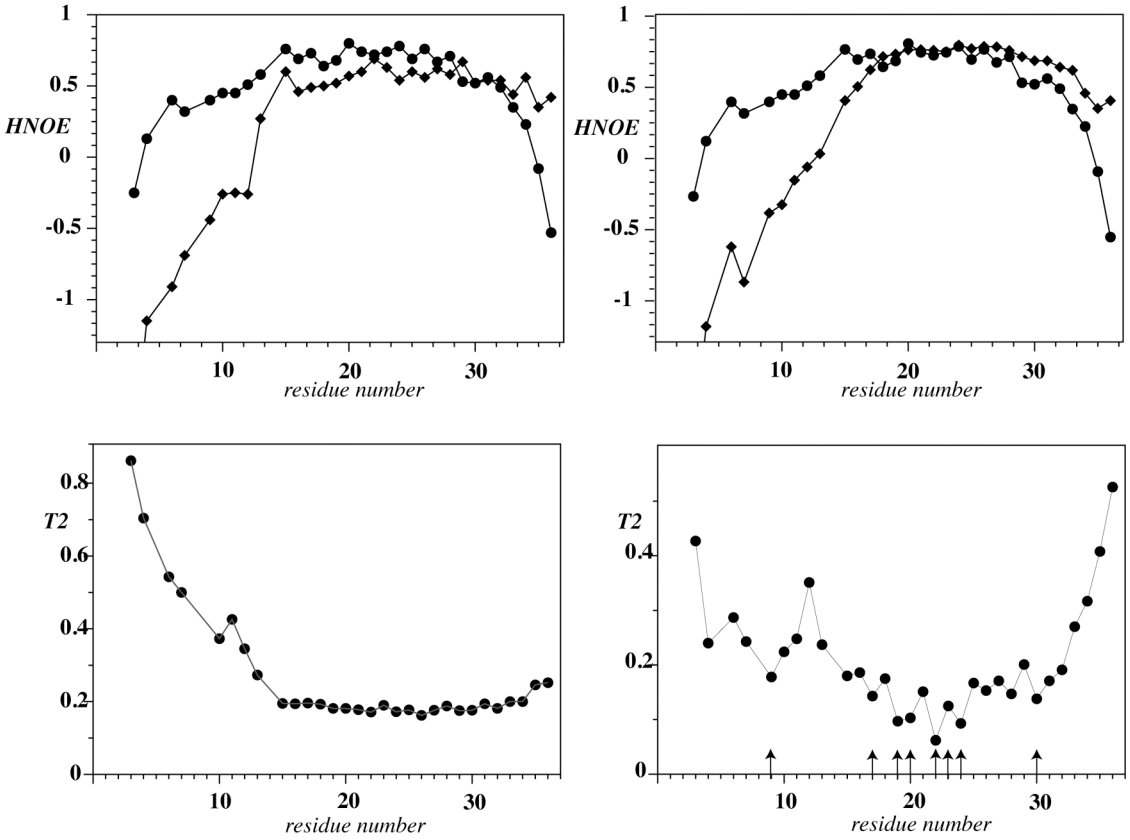


Figure 2: Comparison of $^{15}\{^{1}H\}$-NOE (top) and transverse relaxation times (T2, bottom) of PYY in various environments. Top: Data in water (circles) and in methanol (squares) (left) and in water (circles) and when bound to DPC micelles (squares) (right). Bottom: Transverse relaxation times (*T2*) of amide ^{15}N nuclei at 500 MHz, 301K, 1mM for PYY in methanol (left) and in water (right).

The values of the $^{15}N\{^{1}H\}$-NOE of PYY in pure methanol are depicted in Fig. 2 and support the view that PYY in methanol is not back-folded. The differences in internal

backbone dynamics of PYY in water relative to its micelle-bound state DPC micelles compared to PYY in water vs. PYY in methanol are small indicating that the molecule undergoes similar structural transitions in both cases. The NOE values of PYY in methanol for residues preceding Ser13 are negative and increasing (negative) values towards the N terminus indicate a freely diffusing N-terminal segment. This behavior is also reflected in the *R2* rates for transverse ^{15}N relaxation of the amide ^{15}N nuclei (see Fig. 2): Interestingly, the *R2* values for residues 14 to 31 comprising the core part of the C-terminal α-helix are very uniform in PYY dissolved in methanol. In contrast, *R2* for the same residues of PYY dissolved in aqueous acetate buffer display considerable variability. We have attributed those fluctuations to exchange contributions to *R2* stemming from motions of the N-terminal segments, and this view is supported by the fact that additional exchange terms R_{ex} were required to successfully model the order parameter S^2 in the modelfree analysis for amide moieties of aqueous PYY in that segment. Our *R2* data also seem to indicate that PYY in methanol displays little tendency to self-associate. The overall correlation time was estimated from the *R2*/*R1* ratio of residues within the helix, and resulted in a value for τ_c of 3.90 ± 0.20 ns, slightly less than the 4.52 ± 0.04 ns for PYY in aqueous solution, and the difference may be due to the reduced viscosity of methanol.

Studies of Unfolding

Thermal Unfolding:

The helical hairpin in PYY is stabilized by a number of weak contacts, mainly involving Tyr residues from the C-terminal α-helix and Pro residues from the N-terminal part. We could recently demonstrate that single point mutations of these residues to Ala result in significant destabilization of the tertiary structure of PYY, indicating that formation of tertiary structure in PYY occurs in a cooperative manner. In this work we now look at thermal and unfolding of PYY in a detailed manner.

The classical experiment for folding studies is temperature-induced unfolding/folding and hence we have monitored formation of the helical hairpin by ^{13}C-NMR. The ^{13}C-alpha chemical shifts are conformation-dependent, and their deviations

from random coil chemical shifts have been used to identify elements of regular secondary structure, or to improve convergence of structure calculations in the early phases of NOE assignments (Spera, Ikura et al. 1991; Wishart, Sykes et al. 1991; Wishart and Sykes 1994; Luginbuhl, Szyperski et al. 1995). In contrast to other techniques such as fluorescence-based that monitor the overall behavior of a protein the temperature dependence of the ^{13}C-alpha chemical shifts present a convenient probe *with residue resolution* and hence directly probe for cooperativity. Temperature profiles of $^{13}C\alpha$ chemical shifts have recently been measured by Sadqui et al (Sadqi, Fushman et al. 2006) and used to demonstrate downhill folding for the protein BBL from *E. Coli*, although this work has stirred an intense controversy (Ferguson, Sharpe et al. 2007; Sadqi, Fushman et al. 2007).

The temperature profiles of $C\alpha$ chemical shifts for all residues extracted from constant-time [^{13}C,^{1}H]-HSQC experiments, are depicted in Fig. 3 for all those residues for which they could be extracted over the full temperature range with good fidelity. All curves are sigmoidal, and even more importantly, mono-modal, supporting the view that the backbone of PYY unfolds in highly cooperative manner. In all cases the midpoint of transition is about 315K, indicating a simultaneous loss of structure throughout the whole polypeptide. No indications are present in the data that certain parts of the sequence undergo conformational transitions that are uncoupled from the remainder.

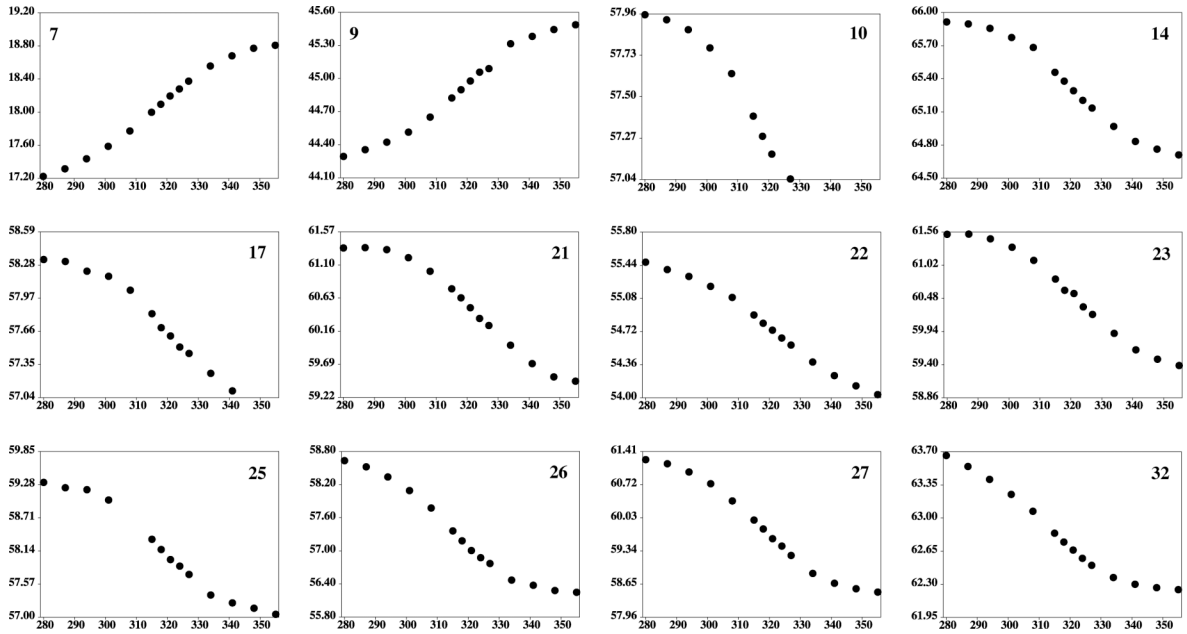


Figure 3: Cα chemical shifts of selected residues of PYY in water at various temperatures with residue numbers annotated in the figures.

Tertiary structure in PYY in water is stabilized by interactions between Tyr residues 20 and 27, located on the hydrophobic side of the C-terminal helix, and Pro residues 2, 5 and 8 in the N-terminal polyproline segment. Loss of tertiary structure may therefore also influence the sidechain conformations of these Tyr residues. Unfortunately, most of the C-beta resonances are highly overlapped, and their chemical shifts cannot be followed. For the isolated C-β resonance of Tyr21 a profile very similar to the profiles from Ca of Tyr21 is observed. (see Fig. 4), and a very similar behavior is also seen for the proton chemical shift difference of the two β-protons of Tyr21. We conclude that the sidechain behavior of Tyr21, a residue that is of prime importance for formation of the helical hairpin, changes its conformational preferences synchronously with the backbone of the polypeptide. This observation is in accordance with changes in proton chemical shifts of the methyl group of Ala-7 and Hδ of Tyr20, presented by us earlier (Neumoin, Mares et al. 2007).

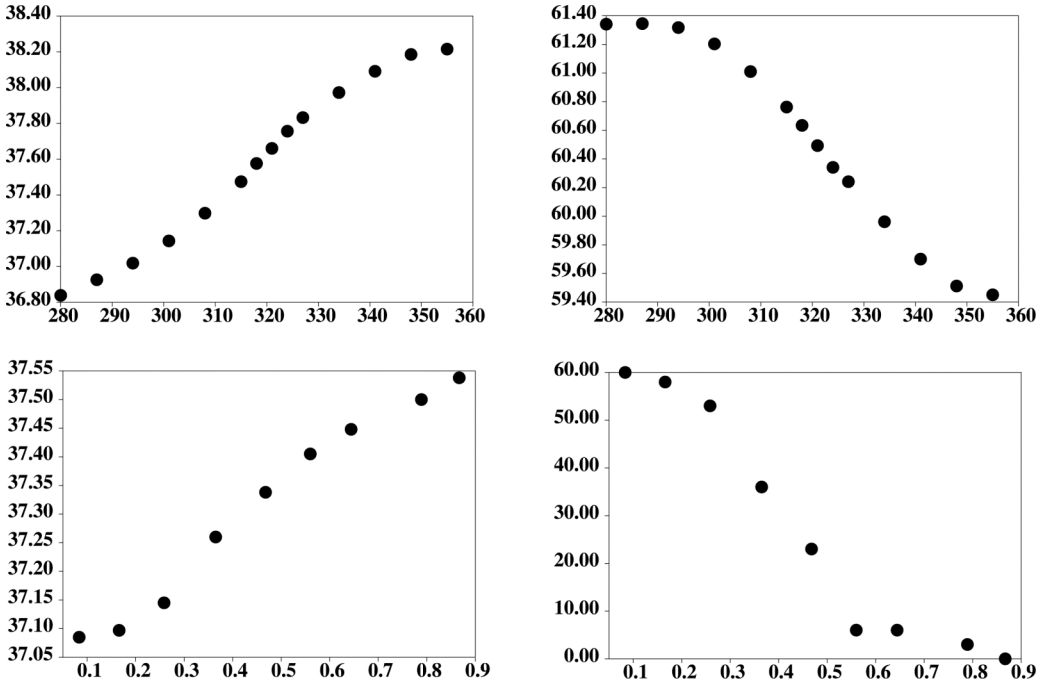


Figure 4: Side-chain chemical shift changes of PYY in heat-induced (top) or solvent-induced (bottom) denaturation experiments. Left: C-β chemical shifts of Tyr21 Right: Difference of proton chemical shifts of the two Hβ-resonances of Tyr21.

Solvent-induced Unfolding:

Removal of contacts between N- and C-terminal residues that establish the helical hairpin is expected to result in complete loss of structure for all residues preceding the first helical residue Pro14. An interesting observation therefore is that the segment comprising residues 12 to 14 apparently is somehow better defined than expected. The N-terminal segment, although largely unstructured, does not fully sample conformational space, and its presence is limited to the space defined by vicinity to the helix and never occurs in the region extending away from the helix. In our previous mutational study we observed that replacement of Ser13 by Ala resulted in loss of tertiary structure, and speculated that the hydroxyl group might be involved in formation of a hydrogen bond that helps stabilizing the conformation of the turn region. In fact, we proposed that the turn, although stabilized by the tertiary contact, is present to some extend independent of the latter. Nevertheless, scalar coupling constants for residues of the turn region are all well-above 6 Hz (see Supplementary Material) indicating that any structure in that region is transient.

Similar to the thermal unfolding studies we have probed the solvent-induced structural transition when the solvent is gradually changed from pure water to pure methanol by monitoring the ^{13}C-alpha chemical shifts. It should be noticed, however, that unfolding in methanol is principally different in nature when compared to thermal unfolding in water: Unfolding in water leads to the complete loss of structure (*vide supra*) while unfolding in methanol results in removal of tertiary structure while secondary structure persists or is even stabilized.

Representative profiles of ^{13}C-alpha chemical shift vs. solvent composition are displayed in Fig. 5. The largest changes in chemical shift ($\Delta\delta > 1$ ppm) are observed in the N-terminal region, particular in the turn region encompassing residues 8 to 13, and in the segment between residues 30 and 33. Interestingly, residues that are part of the turn region display sigmoidal curves indicating that the loss of structure in that region occurs cooperatively. Many other profiles, however, are bimodal and we observe curves with at least two points of inflection indicating that structural changes occur according to a more complex trajectory. This is particularly evident for the profiles of residues Glu16, close to the N-terminal end of the helix, and residues Tyr21 and Ala22, but is also seen for many more residues. We like to emphasize here that in contrast to the thermal denaturation experiment, which lead to the overall loss of both secondary and tertiary structure, the methanol-induced denaturation is more complicated. Firstly, tertiary structure is lost, an event that is nicely observed in the solvent profiles of residues 10 and 14. The removal of tertiary structure is expected to also result in significant destabilization of the C-terminal helix, as was demonstrated by us using single-point mutants of PYY previously (Neumoin, Mares et al. 2007). The structure of PYY in pure methanol, however, has demonstrated that secondary structure in that environment is stabilized at the termini of the helix, and the C-terminal pentapeptide, that is largely flexible in aqueous solution, becomes helically structured. These two counter-acting effects are likely to occur simultaneously and are expected to result in complicated profiles for the Cα chemical shifts *vs*. solvent composition.

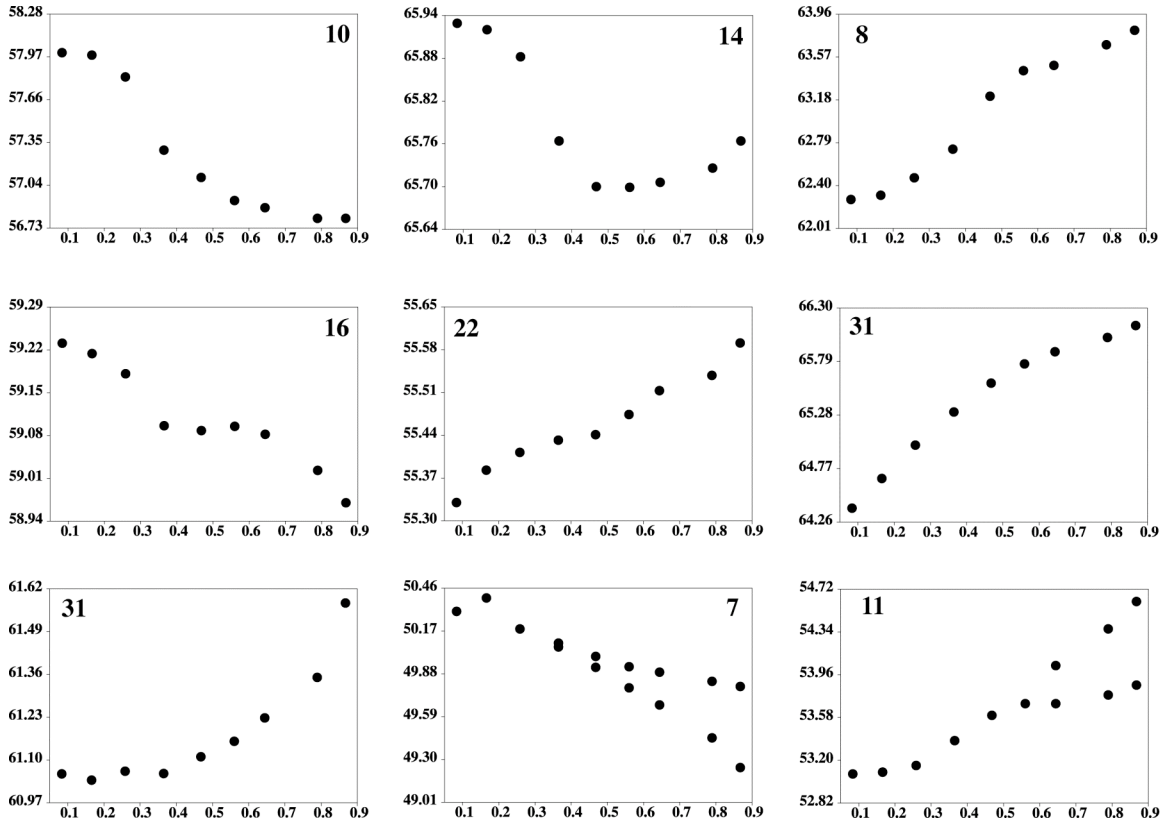


Figure 5: Cα chemical shifts of selected residues of PYY at various water/methanol mixtures (w/w) (1mM, 500 MHz ^{1}H, 301K).

Again, we have followed the isolated C-β resonance of Tyr21 in the various solvent mixtures and obtained a profile very similar to the bimodal curves observed for some of the Cα resonances is seen (see Fig. 4). In addition, the chemical shift difference of the two β-protons of Tyr21 follows a characteristic curve with an inflection point around 35% methanol.

3.3. Discussion

We have recently proposed that the peptide PYY may serve as a mini-protein, and that it posses many properties of globular proteins. The structure of the peptide is surprisingly well defined for such a monomeric short peptide that lacks disulfide bonds or other restraints. Considering its small size data can be recorded at high precision, and observables can be more easily deconvoluted into the underlying effects. Thereby, it becomes a very useful model to study some of the basic concepts of protein folding.

The mutagenesis study performed by us previously indicated that PYY displays high cooperativity in its folding behavior. Using $^{13}C\alpha$ chemical shifts we have shown here that tertiary and secondary structure is destroyed simultaneously in thermal denaturation, whereas secondary structure in methanol-induced denaturation is retained. The dissimilar behavior is due to the differences in the underlying processes of solvent- and thermal-denaturation: In the latter experiments all non-covalent contacts are destroyed, whereas in methanol-induced unfolding experiments only weak hydrophobic contacts are attenuated such that secondary structure is often retained. The comparison of thermal and solvent-induced denaturation experiments therefore allows addressing the importance of hydrophobic contacts for tertiary structure formation experimentally.

The mutagenesis experiments revealed a large role of the turn conformation for adopting the PP fold in PYY, and we proposed that the turn serves as a folding nucleus during hairpin formation. In particular, replacement of Ser13 in PYY by Ala resulted in complete loss of tertiary structure. We speculated that the sidechain of Ser13 may form a hydrogen bond with either backbone or sidechain atoms of Glu15. The now available refined structure of PYY is compatible with an H-bond to the backbone-atom of Glu15, but not to the sidechain amide moiety. The solvent-denaturation profiles of residues 10 and 14 display sigmoidal curves, indicating that structural changes in the turn region occur cooperatively. Apart from the importance of the turn region the solvent unfolding studies also support the importance of multiple weak contacts formed between Tyr 21 and 27 with the N-terminal Pro residues 2,5 and 8. PYY(3-36), a truncated version of PYY, that lacks the Pro2-Tyr27 contact, still forms weak contacts between the N-terminal part and the C-terminal helix (Nygaard, Nielbo et al. 2006), but the hairpin is

significantly destabilized such that tertiary structure was observed only through a very few long-range NOEs. Changes in solvent composition leading to altered lipophilicity of the solvent reduces the magnitude of the favorable Tyr-Pro contacts, and leads to similar changes compared to those arising from the single point Pro-Ala or Tyr-Ala mutations.

Previously, we proposed that some of the destabilized mutants of PYY may serve as models for the molten globular state of PYY. Molten globular forms of polypeptides are characterized by i) structural ensembles that posses overall similarity with the folded state and ii) much more poorly packed sidechains. Mutants of enzymes behaving as molten globules may display almost wild-type biological activity (Vamvaca, Vogeli et al. 2004). Our present data suggest that PYY at certain water-to-methanol mixtures in the 30-40% range may actually also behave like a molten globule. In such an environment it becomes amenable to a more detailed structural study without the requirement for changes in the primary structure of the peptide.

The comparison of the structure of PYY in methanol to that in aqueous solutions of DPC micelles reveals striking similarity in secondary structure. In both environments no tertiary structure is observed, and the C-terminal α-helix extends over the same region. The most apparent difference in conformation is a bent of the helix in the presence of the DPC micelles, which we have attributed to the curved micelle surface. Otherwise, methanol seems to be a good mimic of the environment in a water-membrane interfacial region, and we have proposed the usage of water-methanol mixtures to mimic the transition when the peptide traverses from bulk solution into the interface. It was also realized that the properties near the water-membrane interface are significantly changed with respect to bulk solution and that such an environment may result in partial unfolding (de Jongh, Killian et al. 1992; Bychkova, Dujsekina et al. 1996) similar to the one observed in these water-alcohol mixtures. The membrane-potential leads to a lowered apparent pH in vicinity of the membrane surface. More importantly, there is a sharp decrease in the average dielectric constant. Since denaturing of the proteins seems to depend on average dielectric properties rather than on specific interactions we felt justified to mimic the transition from bulk solution into the water-membrane interface with water-methanol mixtures of varying relative contents.

3.4. Materials and Methods

Materials

$^{15}NH_4Cl$ and ^{13}C-glucose was purchased from Spectra Stable Isotopes (Columbia, USA) and deuterated solvents were received from Cambridge Isotope Laboratories (Andover, Massachusets, USA). The α-amidating monooxygenase used for the enzymatic conversion of Gly was purchased from Unigene. The KSI fusion plasmid was derived from the pET31b(+) vector (Invitrogen). ^{13}C-labelled PYY was expressed as described in more details previously (Bader, Bettio et al. 2001; Neumoin, Mares et al. 2007). After purification by RP-HPLC it was characterized by MS: theoretical mass 4484.6 Da; experimental mass 4483.9 Da.

NMR spectroscopy and structure determination

The structure elucidation of pPYY in solution was performed on an approximately 1mM sample uniformly enriched in ^{15}N and ^{13}C at 28°C, pH 4.2 (uncorrected meter reading) in 20mM deuterated acetate buffer both in 90% H_2O/10% ^2H_2O and 99% ^2H_2O. ^{15}N relaxation data were recorded on 1mM uniformly ^{15}N labeled samples at similar conditions of pH and temperature. Structural data were measured on a Bruker AV-700 and relaxation data were recorded on a Bruker DRX-500. The $^{15}N\{1H\}$-NOEs were recorded at 50.68 MHz. Procedures for spectroscopy, data processing and data evaluation were identical to those described previously. Residual dipolar couplings were measured in stretched polyacrylamide gels using a standard IPAP (^1H,^{15}N) sequence (Tjandra and Bax 1997) or J-modulation experiments (Tjandra and Bax 1997) for measuring $^{13}C\alpha$,Hα or C',Cα couplings. In all cases peak heights were derived from fitted Gaussians in SPARKY (Goddard) and fitted to

$$I(\delta)=\cos(2*\pi* J(C,H)* (T-\delta))$$

in which T denotes the length of the constant-time period (14ms).

Resonances were assigned following the sequence-specific sequential resonance assignment methodology developed by Wüthrich and coworkers (Wüthrich 1986). ^{13}C resonances were assigned from the known proton assignments and transferred with the help of [^{13}C,^{1}H]-ct-HSQC (Vuister and Bax 1992) spectra and (H)CCH (Bax, Clore et al. 1990) experiments. Structures were calculated using upper distance and torsion angle restraints derived from a 120ms 3D ^{13}C-NOESY spectra using the program CYANA (Güntert 2004). After energy minimization using the AMBER94 all-atom force field (Cornell, Cieplak et al. 1995) with explicit water in the program AMBER6 the 20 lowest energy structures were superimposed over backbone atoms of residues 5-7 and 17-31 and deposited in the Research Collaboratory for Structural Bioinformatics protein database under accession code 2OOO. The conformers were analyzed, including calculation of RMSD values, and figures were prepared within the program MOLMOL (Koradi, Billeter et al. 1996).

The data for the $^{15}N\{1H\}$-NOE and R2 of PYY in water and methanol were recorded on 1mM solutions of the peptides using a proton-detected version of the $^{15}N\{1H\}$ steady-state NOE experiment (Noggle and Schirmer 1971) and the Carr-Purcell-Meiboom-Gill (Meiboom and Gill 1958), respectively. The relaxation rate constants *R2* were determined by fitting volume integrals from spectra recorded for various settings of relaxation delays to three parameter single-exponential functions using the Levenberg-Marquardt algorithm (Press, Teukolsky et al. 1992).

3.5. References:

Alonso, D. O. and V. Daggett (1995). "Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water." J. Mol. Biol. 247(3): 501-20.

Babu, K. R., A. Moradian, et al. (2001). "The methanol-induced conformational transitions of beta-lactoglobulin, cytochrome c, and ubiquitin at low pH: a study by electrospray ionization mass spectrometry." J. Am. Soc. Mass. Spectrom. 12(3): 317-28.

Bader, R., A. Bettio, et al. (2001). "Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with Unligated NPY and Implications for Receptor Selection." J. Mol. Biol. 305(2): 307-329.

Bax, A., M. Clore, et al. (1990). "$^{1H-1H}$ Correlation via Isotropic Mixing of 13C Magnetization, a New Three-Dimensional Approach for Assigning 1H and 13C Spectra of 13C Enriched Proteins." J. Magn. Reson. 88: 425-431.

Bianchi, E., R. Rampone, et al. (1970). "The role of aliphatic alcohols on the stability of collagen and tropocollagen." J. Biol. Chem. 245(13): 3341-5.

Blundell, T. L., J. E. Pitts, et al. (1981). "X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone." Proc. Natl. Acad. Sci. U S A 78(7): 4175-4179.

Brutscher, B., R. Bruschweiler, et al. (1997). "Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy." Biochemistry 36(42): 13043-53.

Bychkova, V. E., A. E. Dujsekina, et al. (1996). "Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface." Biochemistry 35(19): 6058-63.

Cornell, W. D., P. Cieplak, et al. (1995). "A Second Generation Force Field For the Simulation of Proteins, Nucleic Acids, and Organic Molecules." J. Am. Chem. Soc. 118(9): 2309.

de Jongh, H. H., J. A. Killian, et al. (1992). "A water-lipid interface induces a highly dynamic folded state in apocytochrome c and cytochrome c, which may represent a common folding intermediate." Biochemistry 31(6): 1636-43.

Ferguson, N., T. D. Sharpe, et al. (2007). "Structural biology: analysis of 'downhill' protein folding." Nature 445(7129): E14-5; discussion E17-8.

Fersht, A. R. and V. Daggett (2002). "Protein folding and unfolding at atomic resolution." Cell 108(4): 573-82.

Goddard, T. D., Kneller, D. G. "SPARKY 3." University of California, San Francisco.

Güntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol. Biol. 278: 353-78.

Harding, M. M., D. H. Williams, et al. (1991). "Characterization of a partially denatured state of a protein by two-dimensional NMR: reduction of the hydrophobic interactions in ubiquitin." Biochemistry 30(12): 3120-8.

Herskovits, T. T., B. Gadegbeku, et al. (1970). "On the structural stability and solvent denaturation of proteins. I. Denaturation by the alcohols and glycols." J. Biol. Chem. 245(10): 2588-98.

Kabsch, W. and C. Sander (1983). "Dictionary of Protein Secondary Strucutre: Pattern Recognition of Hydrogen-Bonded and Geometrical Features." Biopolymers 22: 2577-2637.

Kamatari, Y. O., T. Konno, et al. (1996). "The methanol-induced globular and expanded denatured states of cytochrome c: a study by CD fluorescence, NMR and small-angle X-ray scattering." J. Mol. Biol. 259(3): 512-23.

Keire, D. A., M. Kobayashi, et al. (2000). "Solution structure of monomeric peptide YY supports the functional significance of the PP-fold." Biochemistry 39(32): 9935-42.

Koradi, R., M. Billeter, et al. (1996). "MOLMOL: a program for display and analysis of macromolecular structures." J. Mol. Graph. 14(1): 51-5, 29-32.

Lerch, M., M. Mayrhofer, et al. (2004). "Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors." J. Mol. Biol. 339(5): 1153-68.

Luginbuhl, P., T. Szyperski, et al. (1995). "Statistical Basis For the Use Of C-13-Alpha Chemical Shifts In Protein Structure Determination." J. Magn. Res. Series B 109(2): 229-233.

Meiboom, S. and D. Gill (1958). "Modified spin-echo method for measuring spin-relaxation rates." Rev. Sci. Instrum. 29: 688-691.

Mierke, D. F., H. Durr, et al. (1992). "Neuropeptide Y. Optimized solid-phase synthesis and conformational analysis in trifluoroethanol." Eur. J. of Biochemistry 206(1): 39-48.

Nakano, T. and A. L. Fink (1990). "The folding of staphylococcal nuclease in the presence of methanol or guanidine thiocyanate." J. Biol. Chem. 265(21): 12356-62.

Neumoin, A., J. Mares, et al. (2007). "Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution: Determinants of Stability of a Helical Hairpin." J. Am. Chem. Soc.: .

Noggle, J. H. and R. E. Schirmer (1971). The Nuclear Overhauser Effect - Chemical Applications. New York, Academic Press.

Nygaard, R., S. Nielbo, et al. (2006). "The PP-fold solution structure of human polypeptide YY and human PYY3-36 as determined by NMR." Biochemistry 45(27): 8350-7.

Press, W. H., S. A. Teukolsky, et al. (1992). Numerical Recipes. The Art of Scientific Computing. Cambridge, Cambridge University Press.

Sadqi, M., D. Fushman, et al. (2007). " Structural BiologyAnalysis of 'downhill' protein folding; Analysis of protein-folding cooperativity (Reply)." Nature 445: E17-18.

Sadqi, M., D. Fushman, et al. (2006). "Atom-by-atom analysis of global downhill protein folding." Nature 442(7100): 317-21.

Shiraki, K., K. Nishikawa, et al. (1995). "Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding." J. Mol. Biol. 245(2): 180-94.

Spera, S., M. Ikura, et al. (1991). "Measurement of the exchange rates of rapidly exchanging amide protons: application to the study of calmodulin and its complex with a myosin light chain kinase fragment." J. Biomol. NMR 1(2): 155-65.

Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science 278(5340): 1111-4.

Tjandra, N. and A. Bax (1997). "Measurement of dipolar contributions to 1JCH splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra." J. Magn. Reson. 124(2): 512-5.

Vamvaca, K., B. Vogeli, et al. (2004). "An enzymatic molten globule: Efficient coupling of folding and catalysis." Proc. Natl. Acad. Sci. U S A 101(35): 12860-12864.

Vuister, G. W. and A. Bax (1992). "Resolution Enhancement and Spectral Editing of Uniformly 13C-Enriched Proteins by Homonuclear Broadband 13C Decoupling." J. Mag. Reson. 98: 428.

Wang, Y.-F., M.-Y. Ho, et al. (2004). "Using mass spectrometry to probe subtle differences in conformations of several cytochromes c in aqueous and methanol solutions." J. Mass Spectrom. 39: 1523-1530.

Wishart, D., B. Sykes, et al. (1991). "Relationship between nuclear magnetic resonance chemical shift and protein secondary structure." J. Mol. Biol. 222(2): 311-33.

Wishart, D. S. and B. D. Sykes (1994). "Chemical shifts as a tool for structure determination." Methods Enzymol. 239: 363-92.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. New York, Wiley-Interscience.

3.6. Supplementary material

Table S1: Chemical Shifts of pPYY, 1mM, d^3-MeOH, 301K, referenced to residual $CD_2\mathbf{H}OH$ at 3.31 ppm

N	Res.	H^N	H'^a	H^b	others
1	Tyr	-	-	-, -	δH -, -; εH -, -; ηOH -
2	Pro		4.43	1.97, 2.22	γCH_2 -, -; δCH_2 -, -
3	Ala	8.29	4.32	1.38	
4	Lys	8.16	4.63	1.86, 1.86	γCH_2 1.69, 1.69; δCH_2 1.51, 1.51; εCH_2 -, -; ζNH_3^+ -
5	Pro		4.49	-, -	γCH_2 2.09, 2.09; δCH_2 3.65, 3.80
6	Glu	8.29	4.36	1.94, 1.94	γCH_2 2.11, 2.43; εH -
7	Ala	8.07	4.63	1.34	
8	Pro		4.35	1.97, 2.23	γCH_2 -, -; δCH_2 -, -
9	Gly	8.52	3.80, 3.92		
10	Glu	8.07	4.29	2.42, 2.42	γCH_2 2.07, 2.15; εH -
11	Asp	8.27	4.67	2.79, 2.89	δH -
12	Ala	7.95	4.49	1.39	
13	Ser	8.13	4.7	3.98, 4.26	γOH -
14	Pro		4.3	1.94, 2.05	γCH_2 1.69, 1.69; δCH_2 -, -
15	Glu	8.24	4.09	2.21, 2.21	γCH_2 2.51, 2.59; εH -
16	Glu	7.96	4.11	2.17, 2.28	γCH_2 2.51, 2.51; εH -
17	Leu	8.34	4.11	1.86, 1.86	γH 1.50; δCH_3 0.93, 0.93
18	Ser	8.34	4.25	4.01, 4.10	γOH -
19	Arg	7.97	4.06	1.99, 1.99	γCH_2 1.67, 1.67; δCH_2 3.21, 3.21; εNH 7.45; ηNH_2 -, -
20	Tyr	8.44	4.17	3.13, 3.18	δH 6.76, 6.76; εH 6.56, 6.56; ηOH -
21	Tyr	8.67	4.13	3.12, 3.12	δH 7.16, 7.16; εH 6.77, 6.77; ηOH -
22	Ala	8.58	4.05	1.57	
23	Ser	8.23	4.17	4.03, 4.06	γOH -
24	Leu	8.35	4.1	1.68, 1.68	γH 1.50; δCH_3 0.87, 0.87
25	Arg	8.49	3.96	1.99, 1.99	γCH_2 -, -; δCH_2 3.20, 3.20; εNH 7.44; ηNH_2 -, -
26	His	8.4	4.48	3.41, 3.51	δ^1NH -; δ^2H 7.30; ε^1H 8.73; ε^2NH -

27	Tyr	8.49	4.07	3.21, 3.21	δH 6.97, 6.97; εH 6.64, 6.64; ηOH -
28	Leu	8.7	3.92	1.97, 1.97	γH 1.50; δCH_3 0.93, 0.93
29	Asn	8.55	4.38	2.75, 3.10	δNH_2 7.02, 7.79
30	Leu	8.01	3.97	1.74, 1.74	γH 1.60; δCH_3 0.84, 0.84
31	Val	8.27	3.65	1.98	γCH_3 0.75, 0.84
32	Thr	8.03	3.87	4.36	γCH_3 1.28; γOH -
33	Arg	8.03	4.06	1.98, 1.98	γCH_2 1.86, 1.86; δCH_2 3.16, 3.16; εNH 7.44; ηNH_2 -, -
34	Gln	8.15	4.02	2.27, 2.27	γCH_2 2.07, 2.52; εNH_2 -, -
35	Arg	7.89	4.07	1.65, 1.75	γCH_2 1.43, 1.52; δCH_2 3.01, 3.07; εNH 7.25; ηNH_2 -, -
36	Tyr	7.77	4.52	2.84, 3.20	δH 7.22, 7.22; εH 6.67, 6.67; ηOH -

Figure S2: Sequence plot of restraints used during the structure calculation of pPYY in MeOH

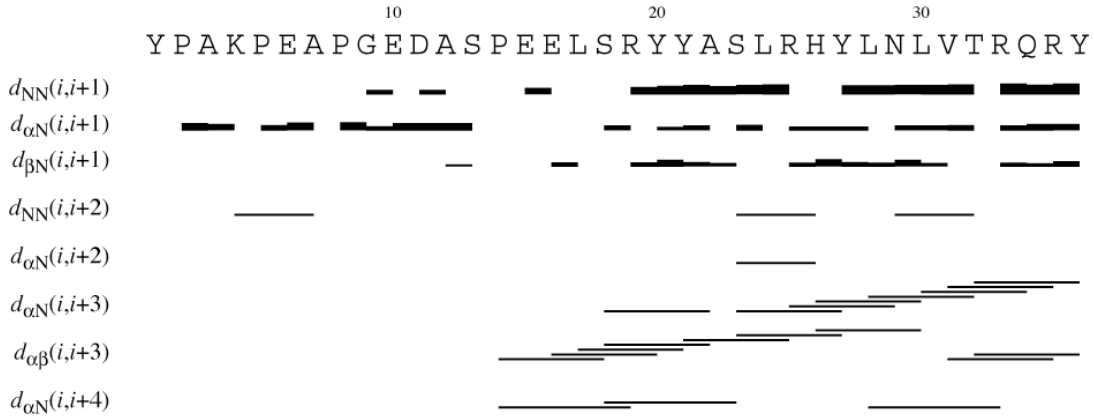


Table S3: Information on the structure calculation of pPYY in MeOH

Distance restraints	Total	229
	Intra-residual	90
	Sequential (i - j = 1)	76
	Medium (i - j = 2, 3, 4)	63
	Long-range	0
Dihedral angle restraints		129
RMSD[a] (Å)[b]	Tyr1-Tyr36 backbone	6.23 ± 1.32
	Tyr1-Tyr36 all heavy atoms	7.63 ± 1.44
	Tyr1-Glu15 backbone	4.82 ± 1.02
	Tyr1-Glu15 all heavy atoms	6.15 ± 1.14
	Pro14-Arg35 backbone	1.73 ± 1.14

	Pro14-Arg35 all heavy atoms	2.72 ± 1.00
	Glu16-Arg35 backbone	1.50 ± 1.00
	Glu16-Arg35 all heavy atoms	2.53 ± 0.88

[a] Atomic root mean square deviation calculated by superimposing backbone atoms of residues 17 to 31 (17 to 31) of the 20 minimized structures referenced to the mean coordinates.

[b] N, Cα, C´ atoms.

Table S7: $^{3}J_{HN\alpha}$ scalar coupling constants for pPYY in MeOH and in water. Values ≤ 6.0 Hz are shaded in grey.

Res.	pPYY in MeOH	pPYY
1		
2		
3	5.9 ± 0.1	6.1 ± 0.2
4	7.1 ± 0.1	5.3 ± 0.3
5		
6	7.7 ± 0.1	7.1 ± 0.2
7	7.6 ± 0.3	5.1 ± 0.3
8		
9	6.1 ± 0.8	7.4 ± 0.7
10	6.2 ± 0.2	5.5 ± 0.3
11	7.1 ± 0.1	9.8 ± 0.1
12	6.6 ± 0.1	5.0 ± 1.3
13	6.4 ± 0.3	5.7 ± 0.4
14		
15	5.7 ± 0.3	5.1 ± 0.3
16	6.1 ± 0.2	6.0 ± 0.5
17	5.3 ± 0.2	5.0 ± 0.4
18	5.0 ± 0.2	4.7 ± 0.5
19	5.2 ± 0.2	6.2 ± 0.2
20	5.0 ± 0.2	5.4 ± 0.8
21	4.8 ± 0.2	5.3 ± 1.0
22	4.9 ± 0.2	4.9 ± 0.3
23	5.1 ± 0.2	6.0 ± 1.4
24	5.0 ± 0.2	6.0 ± 0.2
25	5.0 ± 0.2	5.3 ± 0.9
26	4.9 ± 0.2	5.6 ± 0.2
27	4.9 ± 0.2	6.1 ± 0.4
28	4.7 ± 0.2	4.4 ± 1.2
29	5.0 ± 0.2	5.6 ± 0.4
30	5.1 ± 0.2	6.3 ± 0.3
31	5.0 ± 0.2	6.3 ± 0.1
32	5.1 ± 0.2	6.8 ± 0.3
33	5.2 ± 0.2	6.7 ± 0.3
34	5.3 ± 0.2	6.6 ± 0.2
35	6.0 ± 0.2	6.8 ± 0.2
36	9.3 ± 0.1	8.9 ± 0.2

4. NMR studies in DPC of a fragment containing the seventh transmembrane helix of a GPCR from *Saccharomyces cerevisiae*

The structure and dynamics of a large segment of Ste2p the G-protein coupled α-factor receptor from yeast were studied in dodecylphosphocholine (DPC) micelles using solution NMR spectroscopy. We investigated the 73-residue peptide EL3-TM7-CT40 consisting of the third extracellular loop 3 (EL3), the 7th transmembrane helix (TM7) and 40 residues from the cytosolic C-terminal domain (CT40). The structure reveals the presence of an α-helix in the segment encompassing residues 10 to 30, which is perturbed around the internal Pro24 residue. RMSD values of individually superimposed helical segments 10-20 and 25-30 were 0.91 ± 0.33 Å and 0.76 ± 0.37 Å, respectively. ^{15}N-relaxation and RDC data support a rather stable fold for the TM7 part of EL3-TM7-CT40, whereas the EL3 and CT40 segments are more flexible. Spin-label data indicate that the TM7 helix integrates into DPC micelles, but is flexible around the internal Pro24 site, exposing residues 22 to 26 to solution and reveal a second site of interaction with the micelle within a region comprising residues 43-58, which forms part of a less well-defined nascent helix. These findings are discussed in the light of previous studies in organic-aqueous solvent systems.

published as: Neumoin, A., Arshava, B., Becker, J., Zerbe, O., Naider, F. (2007). Biophys J. 93(2): 467-82

4.1. Introduction

G protein-coupled receptors (GPCRs) constitute a large family of integral membrane proteins of prime biological importance. They are involved in various important physiological processes such as signal transduction associated with cell growth, pain perception, blood pressure control and sensing of light, odour and taste(Mombaerts 1999). Approximately thirty percent of drugs currently used to treat various pathologies target GPCRs.(Lundstrom 2005; Thompson, Burnham et al. 2005) Despite the widespread occurrence of GPCRs and the fact that they have been studied intensively during the last two decades, fundamental information concerning their three-dimensional structure and about the molecular details of ligand binding and signal transduction is still missing. Although more than 1000 GPCRs have been identified, presently only a single high-resolution X-ray structure that for bovine rhodopsin a light-sensing GPCR is available(Palczewski, Kumasaka et al. 2000). The atomic details of the seven transmembrane helical bundle from this crystal structure have served as a scaffold for modeling of other GPCRs (Flower 1999), since all members of this super-family are believed to share a common topology of seven membrane-spanning helices, connected either by extracellular or cytoplasmic loops. The amino and carboxy-termini are always located at the extracellular and cytoplasmic side, respectively (Strader, Fong et al. 1994; Ji, Grossmann et al. 1998; Ballesteros, Shi et al. 2001). The tremendous difficulties encountered in obtaining refraction-grade crystals of membrane proteins, and the large size of their complexes with detergents and lipids complicate structural studies by X-ray crystallography or NMR. Furthermore, expression, purification and reconstitution in a membrane-mimicking environment are technically extremely demanding, and only slow progress has been made despite intense efforts (Sarramegna, Talmont et al. 2003; Grisshammer, White et al. 2005; Sarramegn, Muller et al. 2006).

To address these issues much attention was recently devoted to the study of relatively short peptides corresponding to loops and single transmembrane domains (TMDs) of GPCRs to expand our knowledge on local details of the three-dimensional structure of the intact molecules (Katragadda, Alderfer et al. 2001; Yeagle and Albert 2002). Most of the previous structural investigations on fragments of GPCRs have been

limited to fragments containing up to about 50 residues. Even for these relatively short peptides, only few high-resolution structures in detergent micelles have been published, indicating the practical difficulties encountered in conducting such biophysical studies.

We have performed intensive studies on individual TMDs of the α-factor receptor (Ste2p) from Saccharomyces cerevisiae (Xie, Ding et al. 2000; Naider, Arshava et al. 2001; Valentine, Liu et al. 2001; Arshava, Taran et al. 2002; Naider, Ding et al. 2003; Estephan, Englander et al. 2005; Naider, Khare et al. 2005). Signaling by Ste2p, triggered by binding the tridecapeptide-α-factor mating pheromone, results in growth arrest and gene regulation in preparation for sexual conjugation of yeast cells(Naider and Becker 2004). Like other GPCRs, the 431-residue Ste2p contains seven hydrophobic transmembrane domains (TMDs), with its carboxyl terminus located in the cytosolic milieu. Mutagenesis studies have revealed a key role of the TMDs in α-factor receptor activation. The carboxy-terminus was shown to be involved in Ste2p down-regulation through endocytosis, and in desensitization by phosphorylation(Chen and Konopka 1996).

To determine their structure in hydrophobic environments fragments corresponding to the individual TMDs were studied by CD, IR and NMR spectroscopy in trifluoroethanol (TFE)/water mixtures(Naider, Khare et al. 2005). In three of these domains the α-helices were disrupted by a kink centered around a Pro residue in the case of the sixth and seventh TMDs, and around two Gly residues in the case of first TMD. The solubility of constructs corresponding to the third and fourth TMDs was very low, such that they could not be effectively purified by HPLC, requiring addition of several lysine residues at both termini of the peptides(Melnyk, Partridge et al. 2003). Furthermore, the sixth TMD of Ste2p displayed a high tendency for aggregation on SDS-PAGE(Arshava, Taran et al. 2002). To minimize sample preparation problems and spectroscopic difficulties resulting from poor solubility we envisaged synthesizing constructs of the receptor in which appreciable parts of the hydrophilic cytosolic domain were added to the hydrophobic seventh TMD. (Naider, Ding et al. 2003) In addition to increasing solubility, the cytosolic extension may aid in folding of the construct, and information on its structure could be relevant to its biological role in signal transduction and regulation.

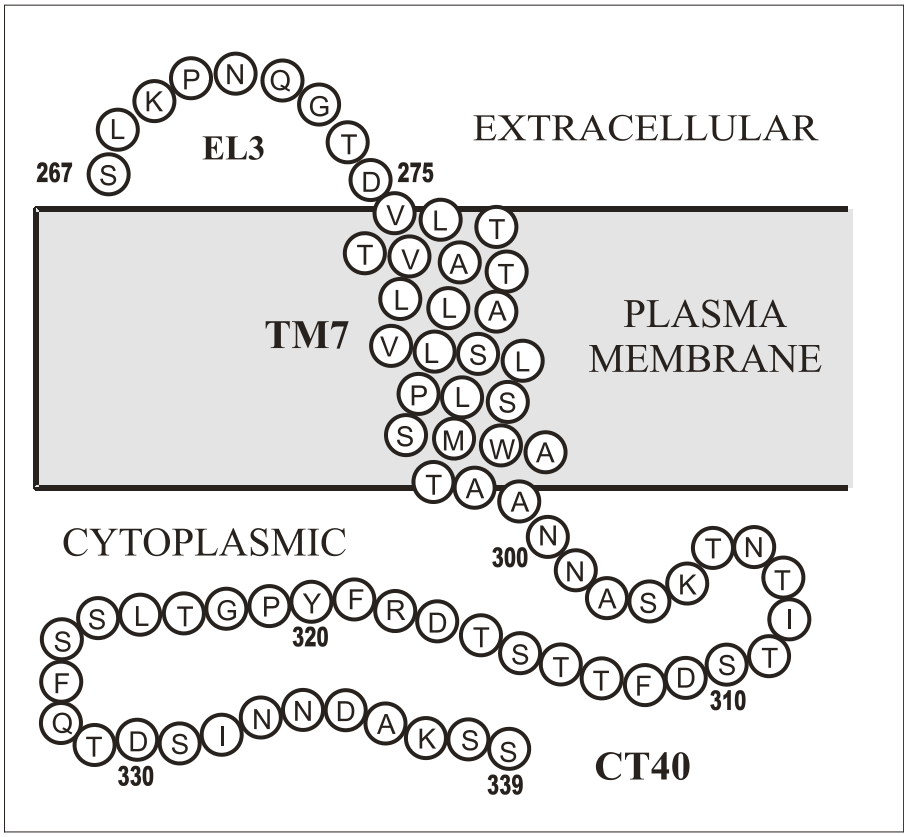


Figure 1: Cartoon of the region of Ste2p examined in this investigation. EL3; Third extracellular loop. TM7; seventh transmembrane helix. CT40; forty residues of the cytosolic tail.

For convenient production of this larger domain we have chosen a recombinant approach in E. coli. The presently described polypeptide is composed of 9 residues of the third extracellular loop (EL3, Ste2p residues 267-275), 24 residues comprising the putative seventh transmembrane domain (TM7, Ste2p residues 276-299) and 40 residues of the cytosolic carboxy-terminus (CT40, Ste2p residues 300-339) of the Ste2p receptor (residues 1-9, 10-33 and 34-73 of EL3-TM7-CT40 respectively, Fig. 1). The desired 73-residue TMD peptide was expressed as a TrpΔLE fusion protein and liberated from its fusion partner using cyanogen bromide cleavage (Estephan, Englander et al. 2005). The recombinant method facilitated expression of EL3-TM7-CT40 in uniformly ^{15}N or ^{15}N, ^{13}C labeled forms required for triple-resonance NMR spectroscopy. Recently we demonstrated that upon addition of ^{15}N-labeled amino acids to rich growth media, peptide selectively labeled with Ala, Ser or Leu residues with acceptable percentages of isotope cross-labeling(Englander, Cohen et al. 2006) were produced.

Herein we report data on the structure and internal backbone dynamics of the multi-domain peptide EL3-TM7-T40 in dodecylphosphocholine (DPC) micelles. The ^{1}H, ^{13}C and ^{15}N resonances could be assigned to a very large extent. The computed structure, based on restraints from NOESY experiments, reveals the presence of an α-helix in the segment 10-30, corresponding to residues 276-296 of the Ste2p receptor, that is disrupted around the internal Pro residue. The internal backbone dynamics derived from ^{15}N relaxation data support the view that the TM part of EL3-TM7-CT40 is rather stably folded, whereas the cytosolic part is much more flexible. Micelle-integrating spin labels support the conclusion that the TM7 helix integrates into DPC micelles, but also demonstrate that motion around the internal Pro site partially brings residues, that would be deeply buried in the micelle interior, closer to the detergent headgroups. The C-terminal decapeptide of the polypeptide is unstructured, but large segments (e.g. from residues 43 to 58) with significant propensity for adopting helical conformations exist. The micelle insertion/association topology of the peptide is fully supported by our understanding of amino acid partitioning into the membrane interior or the membrane-water interface. These studies further validate the approach of using fragments of GPCRs as surrogates to probe receptor structure.

4.2. Results

Sample preparation

Direct dissolution of EL3-TM7-CT40 in DPC solution resulted in poor quality [^{15}N,^{1}H]-HSQC spectra with little signal, indicating that the peptide was not properly inserted into the micelle. Therefore, the protocol developed by Killian et al.(Killian, Trouard et al. 1994) was applied, in which the peptide was dissolved in a 50:50 (v/v) mixture of hexafluoro-i-propanol (HFIP)/water followed by dilution into micellar solution, lyophilization and redissolving in pure water. A sample prepared by this protocol resulted in moderate quality [^{15}N,^{1}H]-HSQC spectra in which approximately 60 out of the 70 expected backbone resonance peaks were visible.

We subsequently modified the protocol to include two steps of lyophilization. The peptide and DPC were initially dissolved in HFIP and lyophilized until an oily residue remained. The latter was taken up in aqueous buffer and thoroughly lyophilized to complete dryness to eliminate residual HFIP. Following this procedure reproducible [^{15}N,^{1}H]-HSQC spectra containing the expected 70 peaks could be measured (Fig. 2), indicating that the EL3-TM7-CT40 was well-integrated into the DPC micelles. Such samples were sufficiently stable for measurement of 3D and 4D NMR spectra at 310K, and only displayed indications of additional peaks after more than 2 weeks. At that time a spectrum with the original quality could be recovered when the sample was lyophilized and redissolved, but after 3-4 weeks additional signals due to degradation appeared.

Resonance assignment

Sequence specific sequential resonance assignment of EL3-TM7-CT40 was accomplished by triple-resonance NMR spectroscopy using uniformly ^{13}C,^{15}N and ^{15}N uniformly labelled samples.

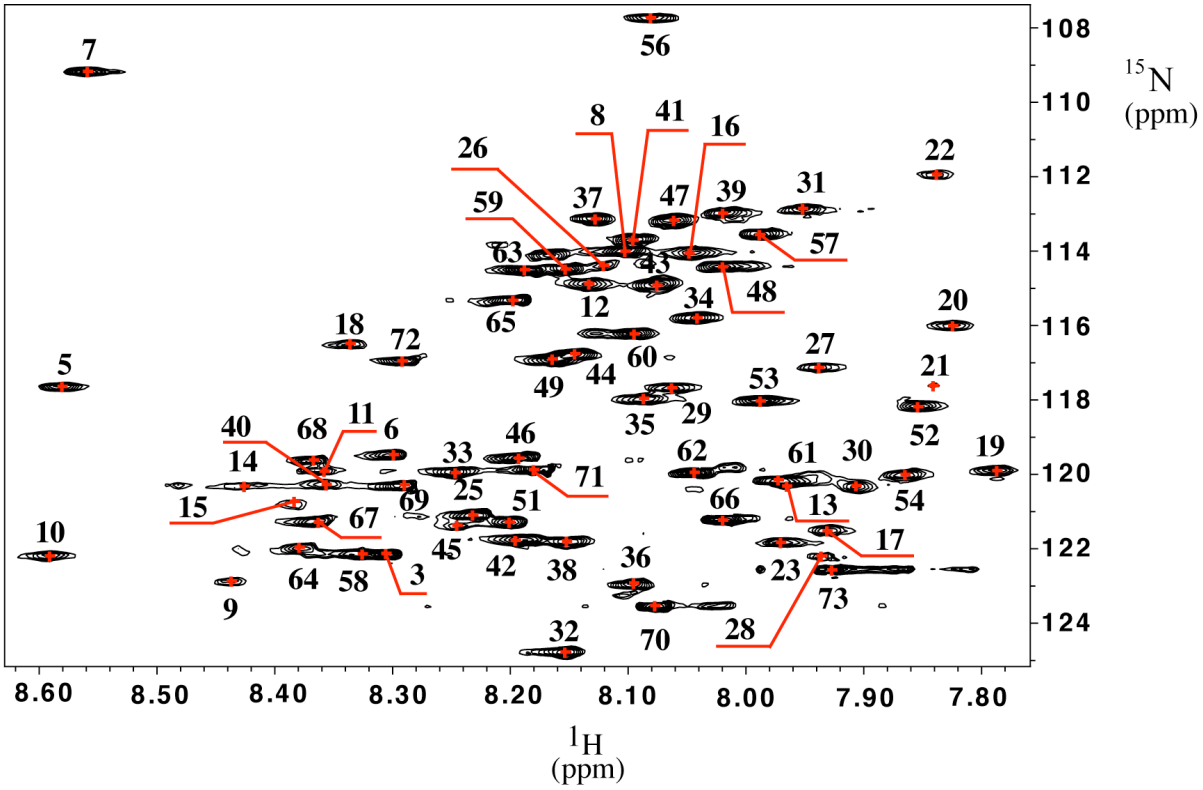


Figure 2: [^{15}N, ^{1}H]-HSQC spectrum recorded at 700 MHz, 310K, on a 0.5mM sample of the peptide in 300 mM DPC. The sequence-specific assignment has been annotated.

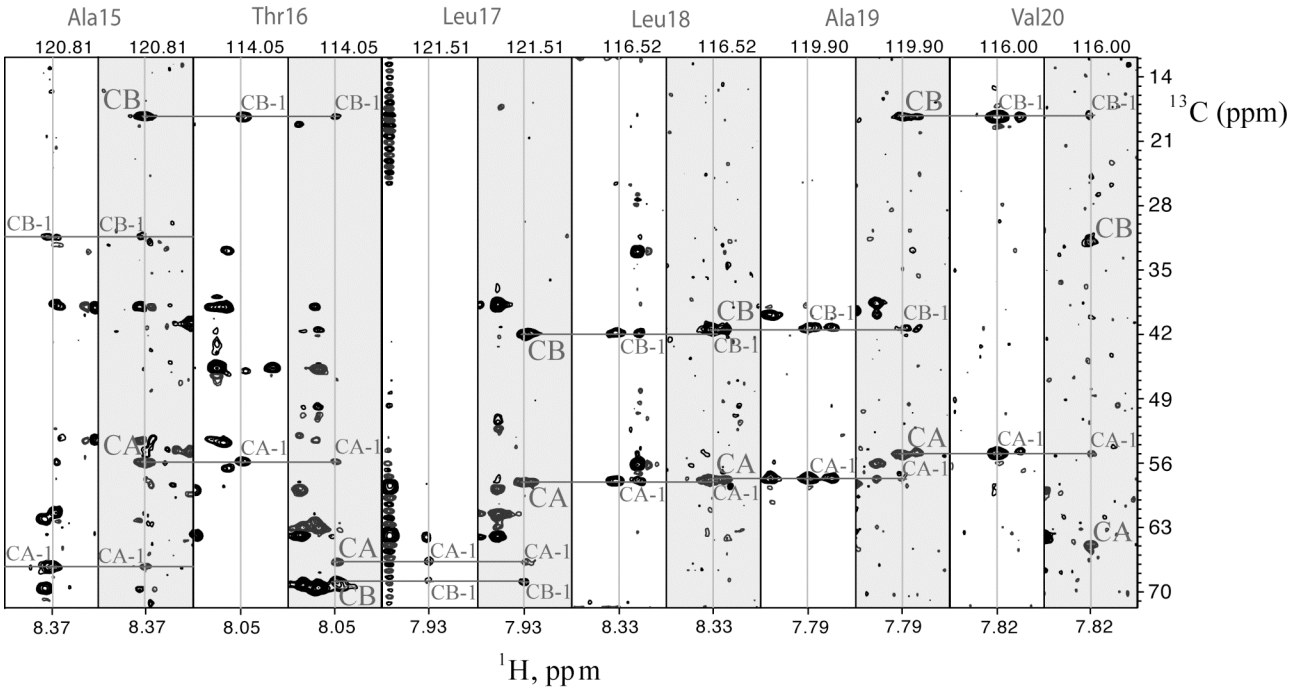


Figure 3: Strips from the 3D CBCA(CO)NH and the HNCACB spectra taking at various amide proton positions displaying the assignment process of the ^{15}N,^{13}C and ^{1}H chemical shifts of backbone and Cβ atoms. The ^{15}N chemical shift, at which the strip was extracted, is displayed above the strips.

Data evaluation was performed using the recently developed program CARA(Keller 2004). For backbone assignment 3D HNCACB(Wittekind and Mueller 1993) and CBCA(CO)NH(Grzesiek and Bax 1992)spectra provided the most useful information. In the [^{15}N, ^{1}H]-HSQC spectra we were able to observe over 60 cross-peaks, for which intra- and/or interresidual C_α and C_β cross-peaks occurred in the HNCACB and CBCA(CO)NH spectra. For the remaining 10 cross-peaks in the [^{15}N, ^{1}H]-HSQC either one or several resonances were missing in the corresponding strips from the HNCACB or CBCA(CO)NH spectra. The unique chemical shifts observed for Gly and Ala residues provided starting points in the assignment process. A set of strips corresponding to the assignment of residues Ala15 to Val20 of the TM segment is depicted in Fig. 3 and the completely assigned [^{15}N, ^{1}H]-HSQC spectrum is shown is Fig. 2. Additionally we recorded [^{15}N, ^{1}H]-HSQC spectra on samples of the peptide that were selectively labeled

by Ala, Ser and Leu (see Fig. 4). These data served to confirm the position of corresponding residues in the [^{15}N, ^{1}H]-HSQC spectra despite some cross-labeling of Ile42 and Ile66 (Fig. 4). Cross peaks from Leu2 and Leu25 could not be found. We experienced significant difficulties in making assignments for residues 21 to 28 and 46 to 51. Due to a lack of correlations in the triple resonance spectra no assignments of ^{1}H and ^{15}N resonances corresponding to Leu2 and Thr50 could be obtained. In order to provide an additional check for resonance assignments, which uses further information from the type of side-chain spin systems, we also recorded an H(CCC)(CO)NH experiment(Montelione 1992).

H_α/C_α cross-peaks from the backbone assignment were subsequently used as anchoring points for further assignment of the aliphatic side chains. Most likely due to short T2 relaxation times of many Cα and Cβ resonances the HCCH-TOCSY spectra(Bax, Clore et al. 1990; Olejniczak, Xu et al. 1992) were unfortunately of moderate quality only, and we were forced to make extensive use of ^{15}N-resolved TOCSY and NOESY and ^{13}C-resolved NOESY spectra to complete the assignment.

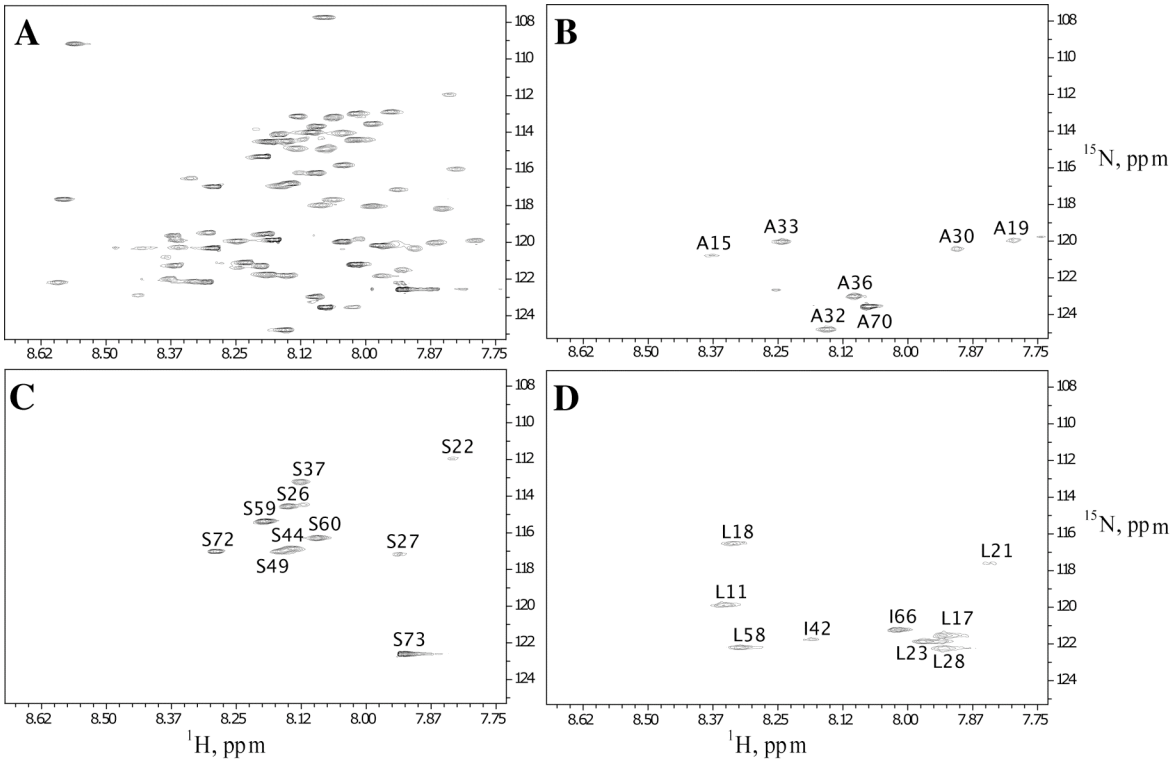


Figure 4: [^{15}N, ^{1}H]-HSQC spectra of selectively labeled constructs. The reference spectrum of ^{15}N-uniformly labeled peptide is shown in panel A. Panels B, C and D display spectra from Ala, Ser and Leu residues, respectively.

In order to accomplish assignment of aromatic side-chains, two-dimensional versions of the (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments(Yamazaki, Formankay et al. 1993) were recorded. Assignments of aromatic protons were completed and verified by data from a [^{1}H,^{1}H]-TOCSY relayed constant-time [^{13}C,^{1}H]-HMQC experiment(Zerbe, Szyperski et al. 1996) and a ^{13}C-resolved NOESY centered at the aromatic carbons. The chemical shifts of all assigned ^{1}H, ^{13}C and ^{15}N nuclei are listed in the supplementary material. The completeness of the backbone and side chain assignments was over 90% for all resonances. Assignment of labile protons from sidechains of Asp, Asn, Gln, Glu, Arg, Lys was generally not possible.

Determination of the structure of EL3-TM7-CT40

The structure of EL3-TM7-CT40 was determined by solution NMR methods using uniformly ^{13}C, ^{15}N or ^{15}N labeled 0.5mM peptide in the presence of 300mM DPC at pH 6.0. The final structure calculations utilized a total of 1018 meaningful NOE upper distance constraints, and 46 angle constraints for the backbone dihedral angles derived from Cα chemical shifts and from $^{3}J_{HN\alpha}$ scalar couplings. A graphical representation of the restraints used in the structure calculation is presented in Fig. 5. Characteristic medium-range $H_{\alpha,\beta}$(i,i+3) NOEs have been observed in the region 10 to 21 and 29-34. The average CYANA target function value obtained was 1.01 ± 0.10 Å indicating the presence of only a few minor violations, the average backbone RMSD to the mean coordinates was 11.80 ± 1.74 Å, and no systematic distance constraint violations remained.

The ensemble of low-energy NMR conformations is depicted in Fig. 6 and displays two helices in the putative TM region of the polypeptide. When individually superimposing backbone atoms from these helical segments the RMSD values are 0.91 ± 0.33 Å and 0.76 ± 0.37 Å for residues 10 to 20 and 25-30, respectively. A calculation of

secondary structure according to the Kabsch-Sander algorithm(Kabsch and Sander 1983) reveals that the first α-helix extends from residues 10 to 22 in 13 of the 20 lowest energy conformers, and of the remaining conformers, 7 possessed a helix C-terminally extended by 1-4 residues. The second α-helix extends from residues 25 to 30 in 9 of the 20 conformers, 5 conformers display helices C-terminally extended by 1-2 residues helix and the remaining ones present less regular helical fragments for residues 25 to 38. Among the low energy conformers one is found with an α-helix encompassing residues 11 to 33, but generally the region 23-27 is poorly defined. Hydrogen bonds between carbonyl oxygen atoms of residue i and amide hydrogen atoms of residue i+4 are observed in more than 50% of the cases in the segments 10 to 22 and 25 to 36.

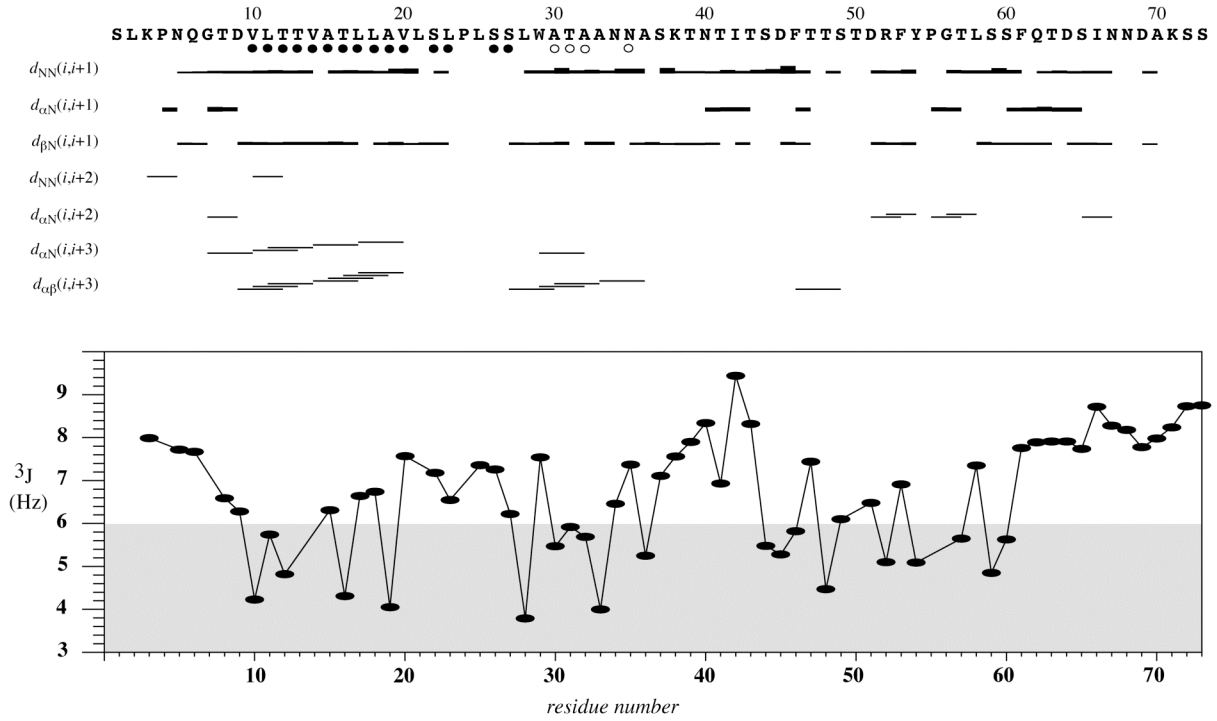


Figure 5: Top: Sequence plot displaying characteristic upper distance restraints along the sequence derived from NOEs. For those residues, for which dihedral angle restraints derived from ^{13}C chemical shifts were applied, a filled circle is placed under the residue number when restraining ϕ in the range of [-120.0.. -20.0°] and ψ to [-120.0.. -20.0°] and for open circles when restraining to the much looser bounds of [-120.0.. 80.0°] for ϕ and [-100.0.. 60.0°] for ψ. Bottom: $^3J_{HN\alpha}$ scalar coupling constants as extracted by the INFIT

method from the HSQC spectrum. The region containing reduced scalar couplings representative of helical tendencies is highlighted in grey.

Considering that residues 23-27 exhibited only weak cross-peaks in the [^{15}N,^{1}H]-HSQC spectrum and that only a few correlations in the corresponding 3D NOESY spectrum were detected, we suspect that this region undergoes slow conformational exchange corresponding to a kink motion of the two helices with respect to each other (vide infra). The helical nature of residues 10 to 20 computed from NOEs are supported by lowered values of the scalar coupling constants $^{3}J_{HN\alpha}$ in this segment of EL3-TM7-CT40 (Fig. 5), and a few couplings below 6 Hz are additionally observed in the segment 27 to 33. In contrast to the putative TM region of the peptide, no elements of regular secondary structure were observed within the CT part throughout all conformers, but 2 conformers possessed backbone dihedral angles corresponding to an α-helix for residues 46 to 48, one conformer exhibited such dihedral angles for residues 55 to 57 and one for residues 60 to 62. We additionally noticed significantly lowered values for the $^{3}J_{HN\alpha}$ coupling constants for most residues in the segment comprising residues 44 to 60. However, we have only observed a few α,N (i,i+2) NOEs and only a single α,β (i,i+3) NOE. The fact that NOEs between sequential amide protons are measured throughout this part of EL3-TM7-CT40, but only a few medium-range NOEs could be detected indicates the presence of transient helical conformations with considerable residual flexibility.

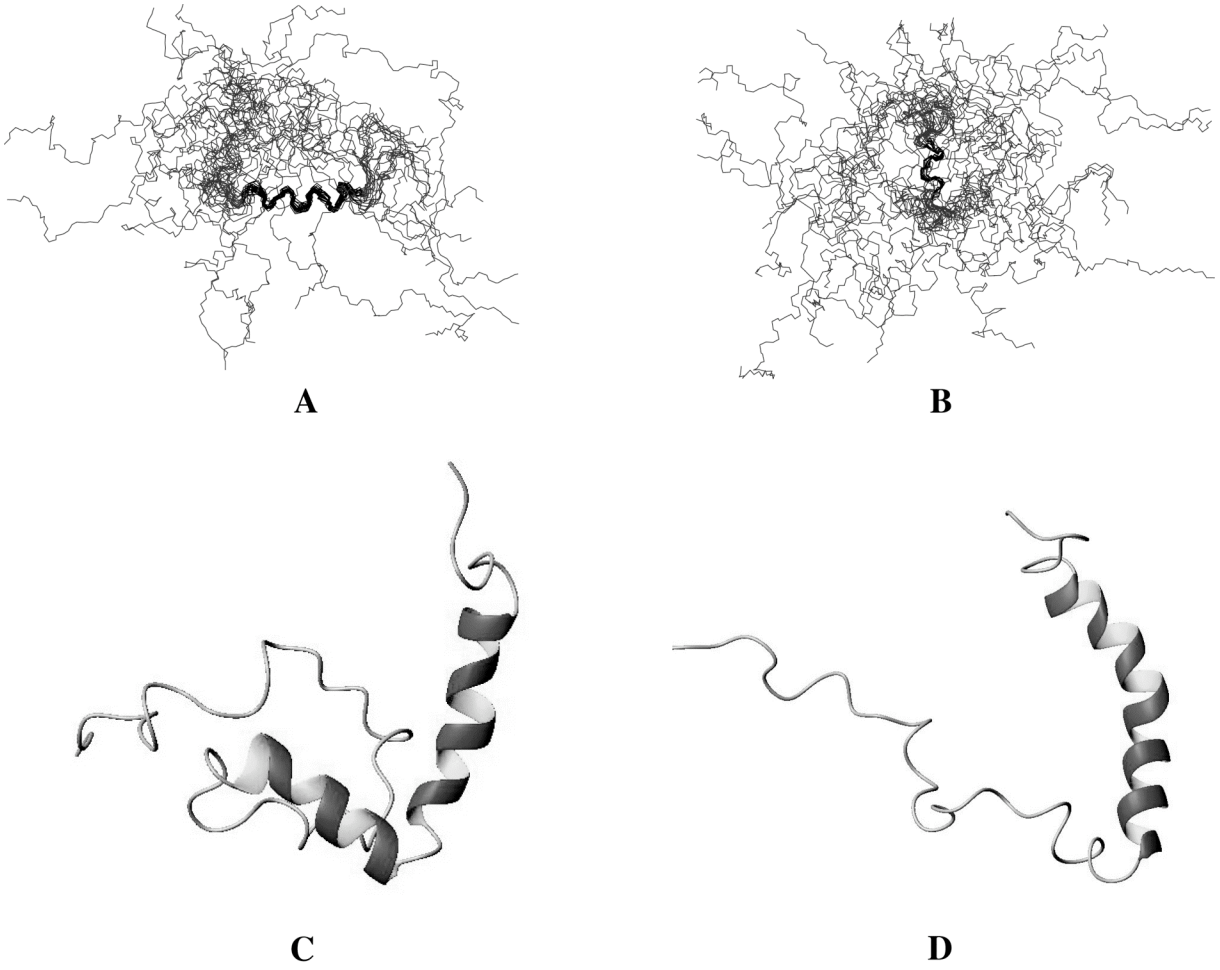


Figure 6: Presentation of the peptide backbone of the 20 lowest energy calculated conformers of E3-TM7-T40 in DPC micelles. Backbone atoms of residues 10-22 (A) and 26-31 (B) have been used to superimpose the structures, and the corresponding bonds are coded in black. Panel (C) and (D) display individual conformers with different dihedrals in the segment comprising Leu23-Pro24-Leu25.

Considering that also the values for the $^{15}N\{^{1}H\}$-NOE are reduced to about 0.4 to 0.55 in that segment (vide infra), we conclude that strong preference for helical conformations exists in that part, but that the persistence of helix conformations is low, similar to what has been referred to in literature as a nascent helix (Dyson, Rance et al. 1988). We suspect that structure in the cytosolic segment is controlled by partitioning of residues into the water-micelle interface (vide infra). When superimposing residues 33 to

73 the RMSD of the backbone atoms is 9.87 ± 2.61 Å indicating high flexibility of this part of the molecule.

In order to further characterize to which extent parts of the polypeptide chain are structured we have measured residual dipolar couplings (RDCs) in stretched polyacrylamide gels. The data are depicted in Fig. 7 and reveal that the RDCs are comparably small, indicating that scaling due to motion occurs. Two segments with increased absolute values (> 1Hz) are observed between residues 10 and 20 and between residues 44 and 53. In other regions the values are below 1 Hz indicating extensive motional averaging.

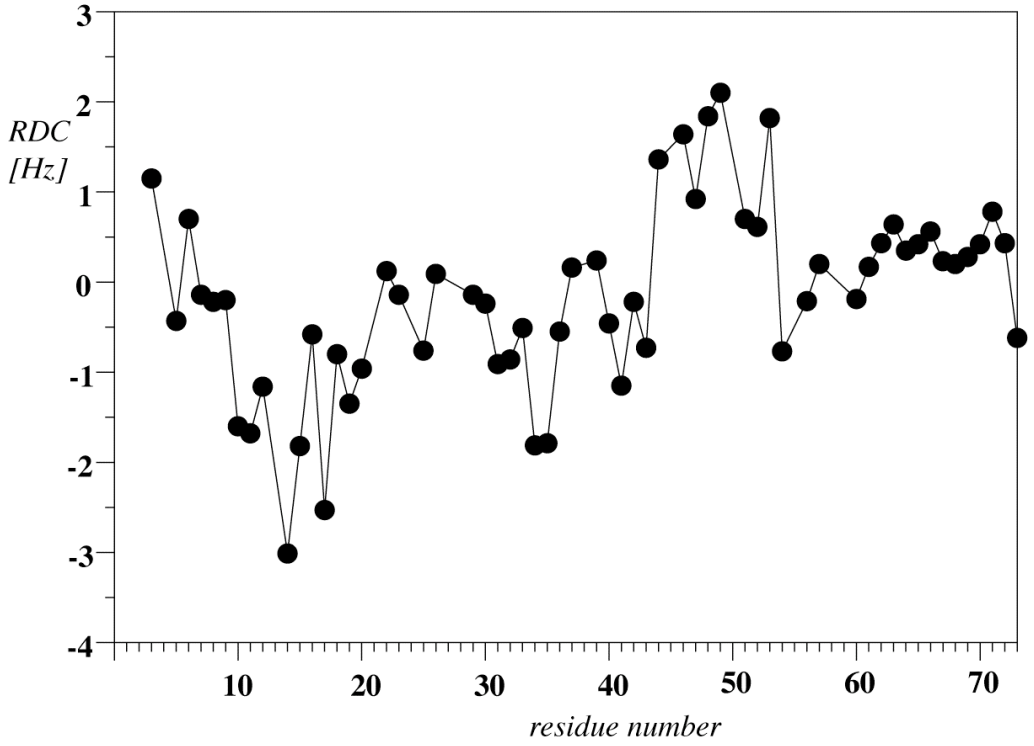


Figure 7: Residual dipolar couplings (RDCs) of E3-TM7-T40 measured at 700 MHz proton frequency.

To further investigate to which extent amide protons were protected from solvent exchange we measured reductions in amide proton intensities due to saturation transfer in an [^{15}N,^{1}H]-HSQC spectrum where the water line was saturated by low-power irradiation during the relaxation delay. Markedly reduced saturation transfer (>70% remaining peak

intensity) was observed in the segment 10 to 30 while rapid exchange was found at the termini of the peptide as well as in the region Asn34 to Thr41 (Fig. 8). Reduced $^{3}J_{HN\alpha}$ and amide proton exchange rates are also observed for residues 44 to 61 supporting the view that this segment is partially structured.

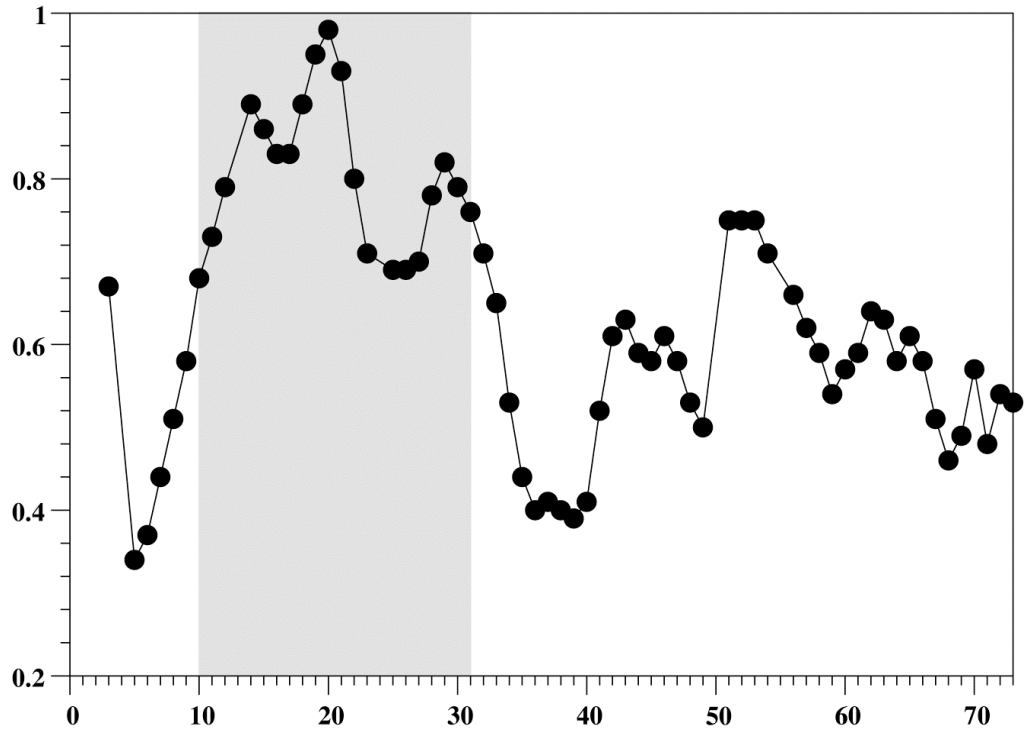


Figure 8: Relative peak volumes of signals computed from HSQC spectra recorded in the presence of low-power presaturation on the water resonance during the relaxation delay relative to a reference experiment without presaturation. In this and in the figures 10 and 11 the putative TM region is shaded in grey.

Dynamics of TM7 as derived from ^{15}N relaxation

It has been generally recognized that less well-defined regions of protein structures computed from NMR solution data may be due to either the intrinsic flexibility of the backbone in that particular segment or to an insufficient number of resolved and assigned NOEs, thereby preventing convergence of structure calculations towards a particular conformation. The determination of internal backbone dynamics in principle can be used to distinguish the two cases. In particular, values of the $^{15}N\{^{1}H\}$-NOE allow

discrimination between well-structured regions of the protein chain from those that display flexibility.

The values of the heteronuclear NOE are depicted in Fig. 9. High values (> 0.6) are usually observed in elements of secondary structure as well as in rather rigid short loops. In TM7 only the segment encompassing residues Val10 to Trp29 fall into this category. The values around Pro24 in this segment are significantly reduced.

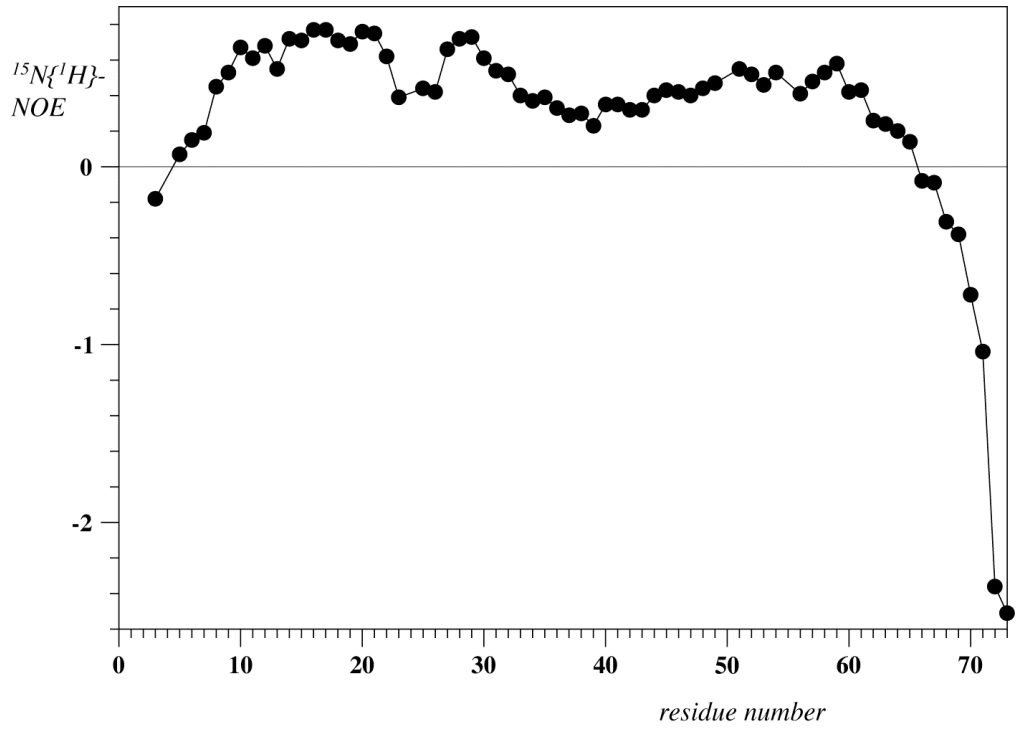


Figure 9: Values of the $^{15}N\{^{1}H\}$-NOE recorded at 700 MHz proton frequency.

A second region of relatively high values NOE (>0.5) values can be found in the segment comprising the putative cytosolic residues Asp51 to Ser59. From residue Gln62 to the C-terminus of the peptide the values of the NOE steadily decrease and adopt (large) negative values towards the termini indicating that the last 10 residues of CT freely diffuses in solution. Residues preceding the putative transmembrane helix (Asp9-Ser1) also exhibit increasingly low values. Thus, based on heteronuclear NOEs both termini are rather flexible.

Orientation and Membrane Integration Topology

Tools from bioinformatics such as secondary structure prediction based on position-specific scoring matrices (Jones 1999; Bryson, McGuffin et al. 2005) predicted the segments Asp9 to Ser22 and Pro24 to Ala32 to be helical (see Supp. Mat.) with high confidence. The analysis of the hydrophobicity profiles concluded that the segment Val10 to Ala33 forms the 7th transmembrane helix in intact Ste2p, but it is unclear whether this segment would adopt such a topology in the presently investigated truncated version of this GPCR in the context of a detergent micelle. Previous fluorescence measurements on the synthetic 64-residue analogue of EL3-TM7-CT40 in the presence of DMPC/DMPG vesicles showed that the single Trp in TM7 was in a hydrophobic environment (Naider, Ding et al. 2003). In order to probe the micelle-integration topology of EL3-TM7-CT40 we determined the effects from paramagnetic relaxation due to the presence of micelle-integrating spin-labels. In this and previous studies we used 5- and 16-doxylstearate, which are presumed to probe the vicinity of the phospholipid headgroups or the membrane interior, respectively. We note that the position of the methyl group at the end of a detergent's aliphatic chain is poorly defined. In contradiction to many pictures found in textbooks a radial extension of the lipid chains from the center of the micelle would result in uneven distributions of atoms across the micelle and, in particular, higher atom density in the center. To account for this fact a statistical model has been proposed by Dill(Dill and Flory 1981) wherein the termini of the lipid chain partially bend back towards the micelle surface, a behaviour that has been verified from MD calculations performed on solvated DPC or SDS (sodium dodecyl sulphate) micelles. This observation complicates the interpretation of spin-label results with 16-doxylstearate and indeed we observe signal attenuations corresponding to the amide moieties thought to be located at the interface as well as the interior of the micelle.

The data derived from the two spin labels are depicted in Fig. 10. For a rigid straight helix traversing the micelle, attenuations due to 16-doxylstearate are expected to be strongest for residues located in the center of the micelle (Fig 10 B), with moderate attenuations for residues at the interface, whereas the effects due to 5-doxylstearate should be largely limited to residues located at the interface (Fig. 10 A). Signal

attenuations for residues 64 to 73 are small and indicate that the C-terminal segment does not interact with the micelle surface. In the following we will consider all residues with attenuations larger than 50% as significantly attenuated. The data reveal such attenuations in the presence of 5-doxylstearate for residues 9 to 11 with the maximum at Val10 and for residues 28 to 30 with the maximum around residue Trp29. A broader third segment with strong attenuations occurs between residues 18 to 23 with maximum attenuations around residue Ser22. For 16-doxylstearate strong attenuations are observed for residues 18 to 23 with signals reduced to less than 20% of their original intensity. In addition, we observed much less reduced signals in the segment Leu25-Ser26, indicating that it is not primarily located in the center of the micelle. Furthermore, there is strong evidence for a second site of interaction with the micelle. Both 5-doxylstearate as well as 16-doxylstearate data indicate very strong attenuations for the segment comprising residues 52-57, centered around residue Phe53. The view that these two sites are making contacts with the micelle surface is supported by reduced amide protons exchange (vide supra). Reductions due to the spin labels although to a much smaller extent also occur for residues Ile42-Thr43 and around residue Thr47. In order to better distinguish attenuation from micelle-surface attached moieties from those buried in the micelle interior we have additionally performed experiments with the soluble spin label Gd-DOTA(Hilty, Wider et al. 2004). In this case the spin label is distributed in solution and should therefore probe for solvent-exposed amide moieties. The data reveal that amide protons in the segment comprising residues 9 to 29 are largely protected from solvent access and also confirm the presence of the second site of solvent protection in the C-terminal part of the polypeptide chain around residue Phe53.

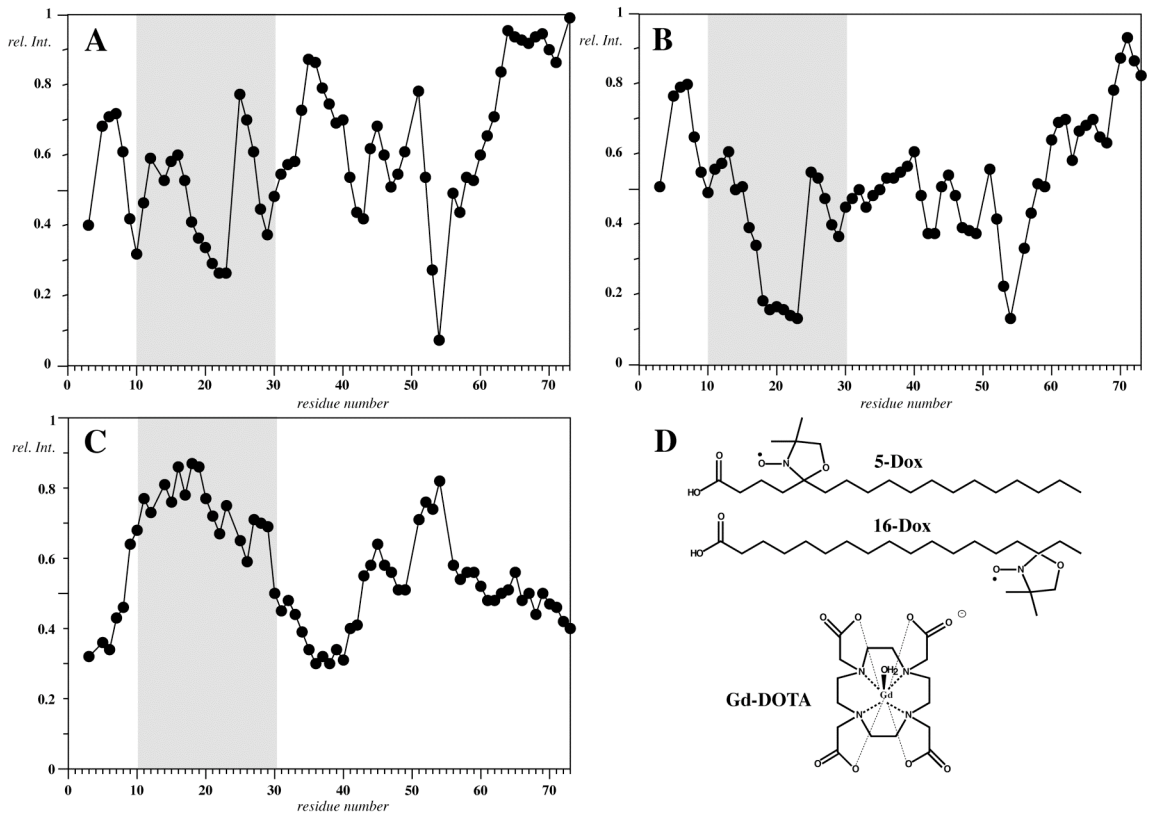


Figure 10: Relative HSQC signal intensities measured on EL3-TM7-CT40 in the presence of various spin labels: 5-doxylstearate (A), 16-doxylstearate (B) and Gd-DOTA (C). The chemical structures of the spin-labels are indicated in (D).

4.3. Discussion

Studies of membrane proteins remain a major challenge in structural biology. Whereas knowledge on beta-barrel membrane proteins is increasing steadily our understanding of structural aspects of helical membrane proteins is still poor. In a remarkable pioneering effort Sanders et al. have recently incorporated the full-length vasopressin GPCR into DPC micelles(Tian, Breyer et al. 2005; Tian, Breyer et al. 2006). However, their TROSY spectra display only 80 of the original 250 peaks indicating strongly reduced *T2* relaxation times for most residues of the GPCR. The authors have proposed that signals from the TM parts most likely are missing. The group of Opella has presented data on full-length CXCR1 in aligned bicelles from solid-state NMR

experiments demonstrating that the receptor was integrated into the bicelles (Park, Prytulla et al. 2006). Despite these promising results with intact GPCRs, direct analysis of these molecules is at present confounded by the immense problems associated with producing full-length, biologically active receptors, purifying and reconstituting these molecules and measuring high quality NMR spectra in membranes. A number of researchers have looked at fragments of GPCRs as surrogates to obtain biophysical data relevant to the intact protein. Pervushin studied peptides derived from the N terminus of bacteriorhodopsin comprising residues 1-71 in organic solvent mixtures consisting of chloroform and methanol and in SDS micelles(Pervushin, Orekhov et al. 1994). Yeagle and coworkers have synthesized peptides corresponding to the cytosolic loops and transmembrane domains of rhodopsin and bacteriorhodopsin, and studied them by NMR, proposing that the structures of these fragments resemble the corresponding regions in the native receptors(Yeagle, Alderfer et al. 1997; Yeagle, Alderfer et al. 1997; Yeagle, Choi et al. 2001; Yeagle and Albert 2002). Pellegrini and coworkers examined a 27-amino acid peptide derived from the third cytosolic loop of the PTH1 receptor in both the linear form, and when cyclized with an octamethylene linker designed to maintain the proposed distance of 12 Å for the loop anchoring points(Pellegrini, Royo et al. 1996). Pelligrini and Mierke additionally studied the extracellular domain of the PTH1 receptor in the presence of DPC micelles(Pellegrini and Mierke 1999). Recently excellent progress has been made in expressing and isotopically labeling regions of the CB2 receptor containing loops and up to two TMs of this GPCR, but a high resolution structure of these 54 residue and 74 residue peptides is not yet available (Zhao, Zheng et al. 2006; Zheng, Zhao et al. 2006).

In previous investigations on EL3-TM7-CT40 of Ste2p, TFE/water (1:1) and chloroform/methanol/water (4:4:1)(Estephan, Englander et al. 2005) were used to mimic the membrane environment. Although the properties of interfacial or core regions of the membrane can be imitated by such solvent mixtures, phospholipid micelles are better mimics of biological membranes, because they possess a completely non-polar interior and a steep gradient of charge density at the water interface similar to that of a bilayer, and allow the N-terminus and CT of the receptor to be exposed to an aqueous environment. Both micelles(Damberg, Jarvet et al. 2001) and the recently introduced

mini-bicelles(Vold, Prosser et al. 1997) have found widespread use in solution NMR studies of peripheral as well as integral membrane peptides/proteins. The Garvin(Krueger-Koplin, Sorgen et al. 2004) and Sanders(Sanders, Hoffmann et al. 2004) laboratories have recently compared detergents for NMR studies of membrane proteins. DPC has been extensively used to study membrane proteins and peptides(55) and taking into account solubilization, stabilization and functional reconstitution of integral membrane proteins, 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG) was concluded to be a detergent of choice for measurement of NMR spectra(54). We measured [^{15}N,^{1}H]-HSQC spectra of EL3-TM7-CT40 in various membrane-mimicking environments including SDS, DPC, 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine and LPPG and observed the best sample stability and good spectral quality in DPC solution.

The present study aimed at elucidating the structure of a fragment from the yeast GPCR Step2p, including the 7th TM domain, the EL3 loop and 40 residues from the CT tail in DPC micelles. Based on the structural information, dynamics and spin-label data we propose the following picture for integration of the peptide into the micelle: The segment encompassing residues Val10 to Ala30 forms a helix that integrates into the interior of the micelle. This α-helix is disrupted around Pro24, and considerable flexibility exists such that the orientations of the N- and C-terminal parts of the α-helix are not well defined with respect to each other.

The hydrophobic core region of biologically relevant membranes was characterized from X-ray and neutron diffraction data on a dioleoylphosphocholine bilayer(Wiener and White 1992) and is approximately 25 to 30 Å in thickness, with the distance between phosphorous atoms adopting values of about 40 Å. In addition, a very steep gradient of charge density exists in the shell located between 10 and 30 Å from the bilayer center. The distance of the phosphorous atoms from the micelle center as extracted from trajectories of MD calculations performed on a 54-lipid DPC micelle aggregate is approximately 17 Å(Tieleman, van der Spoel et al. 2000), compared to 20 Å in the bilayer(Wiener and White 1992). Given this considerable mismatch in hydrophobic thickness of DPC and a bilayer it is reasonable to assume that either the lipid or peptide

would adapt structurally to minimize unfavorable hydrophobic interactions between nonpolar peptide residues and the aqueous buffer.

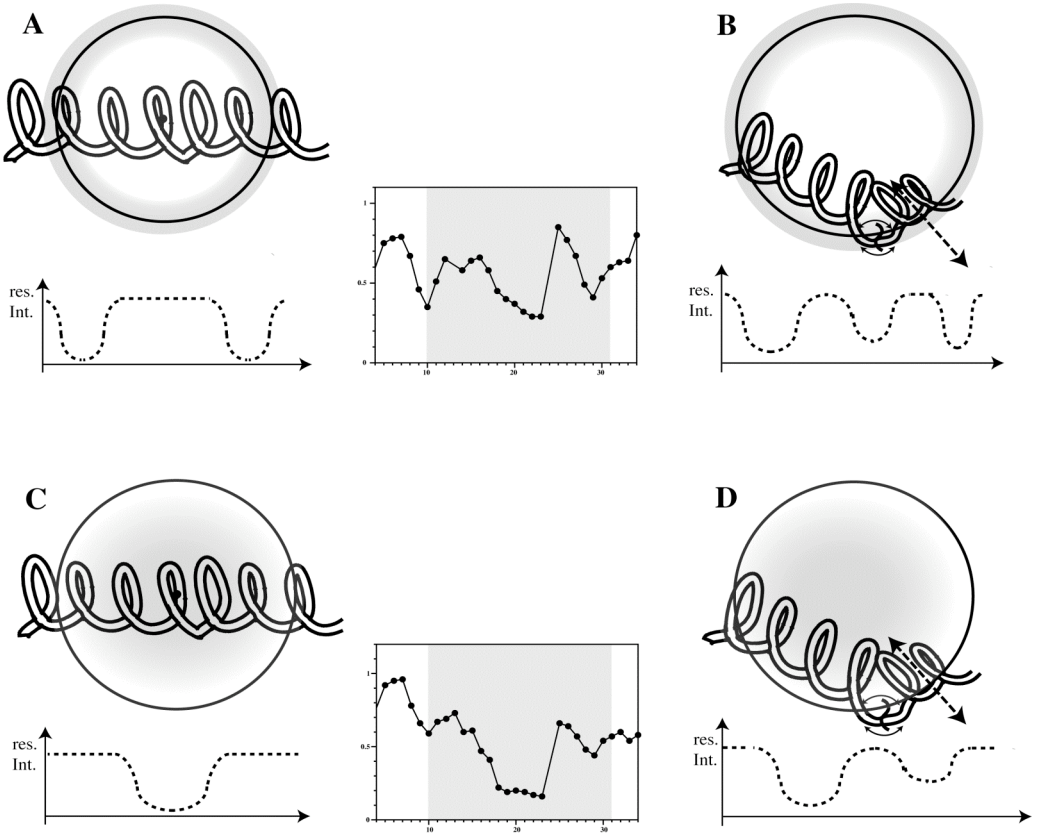


Figure 11: Possible insertion modes of the TM portion of the peptide into the DPC micelle. The areas of largest influence for the two spin-labels 5-doxylstearate (top, A&B) and 16-doxylstearate (bottom, C&D) have been shaded. Below the micelles expected residual signal intensities for the various insertion topologies are depicted. In the center the experimental signal attenuations are shown for a segment containing the putative TM segment (see text).

In the case of the rigidly structured beta-barrel outer membrane protein of *E. coli*, Wüthrich and coworkers could demonstrate that it is the lipids that rearrange such that the hydrophobic part of OmpX is completely covered(Fernandez, Hilty et al. 2002) and data based upon MD calculations of a detergent-OmpX complex suggested a prolate shape of the mixed micelle (Bockmann and Caflisch 2005). In contrast the helical hydrophobic stretch of EL3-TM7-CT40 is much less rigid, and more easily adaptable to the micelle requirements. Our data demonstrate that in case of EL3-TM7-CT40 bending of the helix

helps to accommodate all hydrophobic residues in the 10-30 region of EL3-TM7-CT40 in the interior of the micelle. The situation is schematically sketched in Fig. 11. The positions of the spin-labels 16- and 5-doxylstearate are depicted as gray shaded areas in the center and the outer shell of the micelle, respectively. A straight helix, in principle, should be recognized by the fact, that signal attenuations due to the presence of the different spin-labels are located in different areas of the peptide segment *i.e.* the effects due to 5-doxylstearate would be larger towards the termini (Fig.11A) and the effects due to 16-doxylstearate are expected to be most pronounced in the center of the helix (Fig.11C). In the case of the bent helix, larger parts of the polypeptide segment would actually be located in regions, in which enhanced paramagnetic relaxation due to the presence of 5-doxylstearate occurs (compare Fig. 11 A and B).

In the absence of motion around the helix kink few attenuations for residues in the central segment of the helix due to the 5-doxylstearate spin label are expected, while substantial reductions in the intensities of these residues are expected to occur in the presence of 16-doxylstearate. Motion around the helix kink would be expected to average out differences between the data from the two different spin labels in the central segment of the helix. Moreover, the presence of the polar Ser22, Ser26 and Ser27 residues in proximity to the Pro 24 kink may help to transfer this part of the peptide out of the center of the micelle into the aqueous compartment (notice the reduced attenuation of residues 25-27 from both 5-doxylstearate and 16-doxylstearate in Fig. 10 A and B, respectively). Movement of this part of the peptide to the surface of the micelle would also explain why signal attenuations from the 16-doxylstearate spin label are weaker for residues following Ser26 compared to those in the Val10 to Ser22 segment. We conclude that a significant number of residues from the TM7 helix are integrated into DPC, but considerable motion exists about Pro24. The view presented above is supported by the RDC data. It must be emphasized that the magnitude of the RDCs depends on the orientation of the corresponding NH bond vectors relative to the alignment frame, and it is therefore not directly related to how rigidly a certain segment is folded. However, values larger than 1 Hz are incompatible with extensive motional averaging and hence the data clearly indicate that the N-terminal part of the TM helix is more rigid than the C-terminal part. The lack of significant RDCs, the reduced values of the H-NOE around the Pro-24 site,

and the reduced effects from the spin-labels in the segment 25 to 30 strongly suggest that the C-terminal part of the TM helix is not uniquely oriented relatively to the N-terminal domain (Fig. 7).

The TM helix revealed by the NMR analysis of EL3-TM7-CT40 in DPC is consistent with predictions from bioinformatics (*vide supra*). Surprisingly, the spin label data depicted in Fig. 10 also clearly exposed the presence of a second site of strong interaction with the micelle involving residues centered around Phe53. Both aromatic residues display favorable interaction energies with the micelle-water interface (*vide infra*). The 16-doxylstearate and the Gd-DOTA data (Fig. 10 B and C) as well as the amide proton exchange data indicate that the two sites of interactions with the micelle, the helical region encompassing residues 10 to 30 and the C-terminal region around residue Phe53 integrate differently in the micelle. In particular, the DOTA and the 16-doxylstearate data indicate that the helical segment is buried in the micelle interior over an extended region, whereas interactions with residues Phe53-Tyr54 are limited to a much shorter region and hence can hardly be explained by a large micelle-buried segment. Based on the presence of reduced scalar couplings and the occurrence of NOEs between sequential amide protons we propose that in contrast to the micelle-embedded TM7 the C-terminal part around residue Phe53 is more compatible with the presence of a short surface-associated segment, with a strong preference for helical conformations, which are tightly anchored onto the micelle. Interestingly, the RDCs in this segment have opposite sign to those from the TM helix and therefore indicate that the orientation of the nascent helix in that part is very different from the TM helix. In that respect the RDCs support the view that the cytosolic portion contains a nascent helix, which is surface-associated rather than integrated.

The NMR data that were used for the structure calculation contain few long-range restraints and hence the tertiary structure of the polypeptide seems to be poorly defined. However, the lack of long-range NOEs, reduced RDCs and H-NOEs as well as the effects of spin-labels on different regions of EL3-TM7-CT40 indicate that the lack of tertiary structure is not primarily due to an insufficient number of restraints during the structure calculation. Rather, it likely reflects the flexible nature of this protein, both in the TM helix as well as in the cytosolic part. Recently, the structural role of Pro in TM helices

has been systematically investigated and Pro residues were found to induce kink-motion about the Pro position thereby decoupling the motions of the segment preceding and following the Pro residues(Bright and Sansom 2003).

The topology of membrane association/insertion was recently successfully predicted using experimental thermodynamic parameter for transferring whole amino acids from bulk water into the membrane interface or into the membrane interior as determined by Wimley and White(Wimley and White 1996; White and Wimley 1998). Their data have recently been verified in a biological system using a clever readout system(Hessa, Kim et al. 2005). We have seen that these data reliably predict the orientation of membrane-associated peptides from the NPY family{Bader, 2005 #11)We have seen that these data reliably predict the orientation of membrane-associated peptides from the NPY family (Bader and Zerbe 2005). In Fig. 12 values for transfer into the interior (left) or interface (right) are plotted along the sequence. The segment presenting the transmembrane helix including residues 10 to 30 is immediately recognized because no residues with strongly unfavorable energies for partitioning into the membrane interior are found in that stretch. Moreover, the amphiphilic nature of the C-terminal half of the peptide is obvious with highly hydrophilic residues such as Asp (ΔG_{oct} 3.64 kcal/mol; ΔG_{wif} 1.23 kcal/mol) frequently occurring in the vicinity of hydrophobic residues like Phe (ΔG_{oct} -1.71 kcal/mol; ΔG_{wif} -1.13 kcal/mol). The importance of aromatic residues, in particular of Tyr and Trp, for anchoring polypeptide stretches at the interface has been widely recognized; in fact these residues are highly enriched in interfacial regions (aromatic belt). No such residues are found in the TM region except for Trp29, which likely helps to anchor one end of the TM helix to the interface.

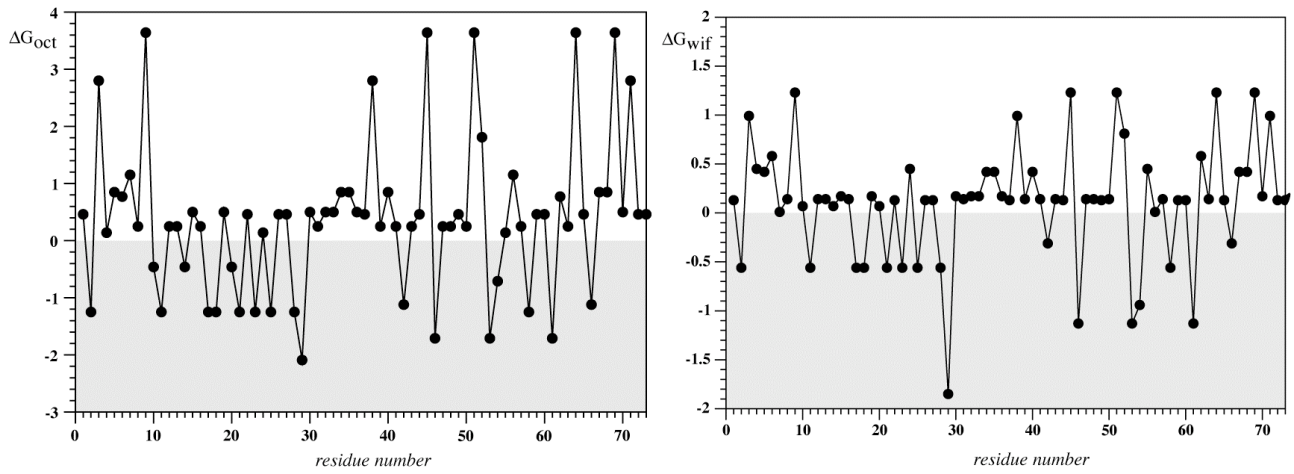


Figure 12: Plot of the free energies for transferring whole amino acids as determined by Wimley and White into the membrane interior (left) and the membrane-water interface (right) for the sequence of the investigated peptide. The area of favorable energies is shaded.

In contrast, four aromatic residues are located in the stretch Phe46-Phe61. We note that the values for the $^{15}N\{^{1}H\}$-NOE are slightly increased for residues 50-60, and these same residues also show more intense NH_i to NH_{i+1} NOEs. We suspect that the stabilization of secondary structure in this region is largely due to anchoring of this region of the peptide chain on the micelle surface. In contrast to the aromatic containing central region of CT40, the C-terminal decapeptide stretch following residue 60 lacks any aromatic residues which could possibly serve as membrane anchors, and displays quite small 5-doxylstearate attenuations and is fully flexible based on dynamics data and the absence of medium- or long-range NOEs.

A systematic analysis of peptide fragments in different solvents and detergents should provide experimental evidence for the specific role played by these media in determining the structure of the polypeptide. Our present work on the structure of EL3-TM7-CT40 in DPC micelles points to both similarities and certain differences in comparison to the organic solvent mixtures. In DPC micelles and organic-aqueous solvents this peptide exhibits a helix encompassing residues 10 to 30 with a kink around Pro24 that results in significant flexibility in both cases. Moreover, certain helical tendencies are revealed for residues in the 43 to 58 range of the peptide both in DPC

micelles and organic-aqueous solvents. However, the C terminus following residue 62 is clearly unstructured in DPC micelles, whereas the C terminal helix in the TFE-water (1:1) extends up to residue 70. Moreover, while the segment 34 to 41, in TFE/water is part of the second helix, it is unstructured in DPC, a conclusion that is supported by the lack of medium range (i,i+3) NOEs as well as NOEs involving sequential amide protons, low values of the $^{15}N\{^{1}H\}$-NOE and little protection from amide proton exchange. In the isotropic environment of an organic-aqueous solvent, structuring effects due to the steep gradient in hydrophobicity in the micelles are absent, which might explain why the cytosolic tail of EL3-TM7-CT40 behaves somewhat differently in organic-aqueous mixtures and in DPC micelles. Thus, it seems that DPC micelles may be a better environment for learning about conformational preferences of the extra and intracellular domains of polytopic molecules.

Taking into account the enormous problems associated with the expression, purification and reconstitution of intact GPCRs into membrane mimetic environments suitable for biophysical studies, investigations on *fragments* of these proteins seem to be well justified. As the size of these fragments increases to include more than one helix our understanding of the involvement of helix-helix interactions in influencing the secondary and tertiary structure of individual TM helices will improve. While many synthetic problems may be reduced when using fragments of GPCRs a crucial question remains whether these truncated constructs are able to successfully mimic structural features of the much longer polypeptides. Many groups have demonstrated that smaller fragments of soluble, well-structured proteins display increased propensity to transiently adopt conformations similar to those encountered for that particular stretch when placed in the context of the full polypeptide (for example see Dyson et al. (Dyson, Rance et al. 1988)). However, these studies have usually also revealed that the fragments are still fairly flexible. This is by no means surprising considering that many crucial and stabilizing medium- or long-range interactions are missing. Moreover, solvent access is not restricted for soluble fragments, and therefore solvation competes with intramolecular hydrogen-bonding. Solvation is much less favorable when the polypeptide is partitioned into a membrane(Popot and Engelman 2000). Indeed, many peptides, which are unfolded in water, adopt secondary structure when placed into a membrane-mimicking

environment, known as the coupled partitioning-folding(White and Wimley 1999). Furthermore, many relatively short peptides fold into stable helices in micelles, both in transmembrane (e.g.(Papavoine, Aelen et al. 1995; MacKenzie, Prestegard et al. 1997; Popot and Engelman 2000)) or surface-associated fashion (e.g.(Brown and Wuthrich 1981; Ladokhin and White 1999; Bader, Bettio et al. 2001)).

Our study has demonstrated that a 73-residue polypeptide comprising the entire 7th TM of Ste2p does adopt a helical conformation even when Pro is part of the helix-spanning stretch. The study, however, has also demonstrated, that the helix is significantly destabilized around Pro24, and that the orientation is presumably not such that the helix is predominantly straight, but rather undergoes larger kink motions. It is possible that such motions are completely or partially suppressed in the presence of the other transmembrane helices of Ste2p, some of which pack against the 7th helix. However, the kink in the structure would be expected to confer residual conformational flexibility that may be very important during signal transduction through Ste2p. The study also indicates that most of the residues in the predicted TM helix have been selected for favorable partitioning into the corresponding membrane compartment while a few have been selected to form crucial helix-helix interactions. We suspect that Ser22 and Ser26 are involved in forming such interactions, because in the absence of other TM segments they tend to promote partitioning of that part of TM7 into the vicinity of the micellar interface. Finally, even in the context of a DPC micelle regions of the CT tail have some propensity to assume transient helical structures. Similar to what we found for the CT of Ste2p, the crystal structure of rhodopsin revealed that the cytoplasmic extension proximal to TM7 contained a helical segment called H8 (Palczewski, Kumasaka et al. 2000). Model peptides corresponding to H8 were studied under a variety of conditions with the conclusion that H8 acts as a membrane-surface recognition domain, where amino acid side chains can interact with phospholipid headgroups (Krishna, Menon et al. 2002). The participation of these "helical" domains in protein-protein interactions with regulatory elements of the signal transduction system remains to be demonstrated.

Conclusions

The present study has demonstrated that polypeptides corresponding to fragments of GPCRs can be incorporated into phospholipid micelles, provided that certain protocols for incorporation are followed, and that detailed information concerning the structure of the peptide and the topology of various regions in the micelle can be deduced. The work revealed details of the folding of a fragment from the yeast Ste2p receptor, and demonstrated that certain structural and/or dynamical features of such a fragment are different in organic solvents and in the presence of DPC micelles. Many features in the latter environment can be explained by anisotropic properties present in micelles and membranes but not in organic-aqueous solvents. The study also demonstrated that even in DPC micelles the isolated 7^{th} TM helix is not rigid and that this flexibility while possibly reduced in the context of TM-TM contacts that exist in the receptor may be an important aspect of the conformational change that is triggered by binding of α-factor to Ste2p. Future studies will be devoted to developing systems that allow study of helix-helix interactions. The present work provides an important starting point for such investigations.

4.4. Materials and Methods

NMR sample preparation

Perdeuterated d_{38}-DPC, perdeuterated MES and deuterated water were purchased from CAMBRIDGE ISOTOPES. The spin labels 5- and 16-doxylstearates were obtained from SIGMA-ALDRICH and Gd-DOTA (DOTAREM™) from LABORATOIRE GUERBET. All other chemical used were ordered from FLUKA.

For preparation of the 0.5mM EL3-TM7-CT40 NMR sample 26.4mg d_{38}-DPC and 1mg of peptide were dissolved together in approximately 1 ml of hexafluoro-*i*-propanol (HFIP), the solution was sonicated for 10 min at 50°C and subsequently lyophilized until an oily residue (12-15hrs) remained. This mixture was dissolved in 250µl of 20mM MES (2-morpholinoethanesulfonic acid, monohydrate) buffer (pH ~ 6),

750µl of water were added and the solution was lyophilized until dryness again. The lyophilized powder consisting of DPC and the peptide was dissolved in 250µl H_2O:D_2O (9:1), mixed by vortexing until all of the solid material is dissolved, incubated for 15 minutes at 37 ºC and transferred to a shigemi NMR tube. The final concentration of d_{38}-DPC was always 300mM. The E3-M7-T40 concentration used for the ^{13}C and ^{15}N-resolved NOESY spectra was 0.5mM, for supporting experiments (e.g. relaxation, spin-label studies etc.) concentrations of 0.1-0.4mM were used. Samples used for the residual dipolar couplings (RDCs) measurements contained EL3-TM7-CT40 and d_{38}-DPC at concentrations of 0.25mM and 200mM, respectively. All NMR measurements were conducted on a Bruker AV700 spectrometer at 310K using a triple-resonance cryoprobe.

For the RDC measurements the peptide-micelle complex was oriented in a stretched polyacrylamide gel (Tjandra and Bax 1997; Sass, Musco et al. 2000). The gel was polymerized from a 4% (w/v) solution of acrylamide and bisacrylamide with a monomer to cross-linker ratio of 37.5:1 (w/w). The dry gel was soaked for 24 h in plain buffer followed by equilibration with a solution of ^{15}N-labelled EL3-TM7-CT40/d^{38}-DPC for 48 h, after which the gel was compressed from 6mm to a final diameter of 4 mm.

Spin label experiments were performed using 5-doxyl stearic acid, 16-doxyl stearic acid or Gd-DOTA in separate experiments. Small aliquots of concentrated solutions of 5- or 16-doxylstearate in d_3-methanol were dissolved in the solution of the ^{15}N-E3-M7-T40/d_{38}-DPC sample to obtain a final concentration of approximately 7mM of spin label corresponding to slightly more than one spin-label per micelle. In case of the experiment utilizing Gd-DOTA an appropriate volume of a 5mM aqueous solution of Gd-DOTA was lyophilized and the remaining powder mixed with the detergent solution of the peptide resulting in a Gd-DOTA concentration of about 6 mM. [^{15}N,^{1}H]–HSQC spectra in the presence and absence of spin labels were recorded and attenuations were computed from the relative peak volumes in these experiments. In cases of overlap peak intensities were used instead of peak volumes and severely or completely overlapped residues were generally excluded from the analysis. In all cases, the relative intensity of residue i was computed from the average of residues i-1, i and i+1, whenever this was possible, to reduce the extent of smaller fluctuations.

Cloning, expression and purification of $^{15}N,^{13}C$-labeled EL3-TM7-CT40

The cloning, expression, and isolation of [$^{13}C/^{15}N$]-EL3-TM7-CT40 {S^{1}LKPN QGTDV L^{11}TTVA TLLAV L^{21}SLPL SSLWA T^{31}AANN ASKTN T^{41}ITSD FTTST D^{51}RFYP GTLSS F^{61}QTDS INNDA K^{71}SS} were carried out using procedures described in the literature (Estephan, Englander et al. 2005). EL3-TM7-CT40 selectively labelled with [^{15}N]-alanine, [^{15}N]-leucine or [^{15}N]-serine was prepared in rich medium containing excess of the [^{15}N]-labelled amino acid as described by Englander et al (Englander, Cohen et al. 2006).

NMR spectroscopy

Sequence-specific resonance assignment was accomplished based on a set of ^{15}N-resolved proton-proton correlation spectra, e.g. a 40ms ^{15}N-resolved TOCSY and 75ms ^{15}N-resolved NOESY(Fesik and Zuiderweg 1990; Clore and Gronenborn 1991), as well as a set of triple-resonance experiments. Backbone assignment was performed based on a CBCA(CO)NH(Grzesiek and Bax 1992) (1024*40(^{15}N)*128(^{13}C) complex data points; t3max 122ms, t2max 12.8ms, t1max 6.1ms), a HNCACB experiment(Wittekind and Mueller 1993) (1024*64(^{15}N)*128(^{13}C) complex data points; t3max 122ms, t2max 20.5, t1max 6.1ms), and a H(CCC)(CO)NH(Montelione 1992) experiment (1024*20(^{15}N)*64(^{13}C) complex data points; t3max 127ms, t2max 10.9, t1max 8.0ms). Sidechain resonances were assigned based on HCCH-TOCSY experiments (1024*32*64 complex points, t3max 104ms, t2max 7.9ms, t1max 9.1ms)(Bax, Clore et al. 1990; Olejniczak, Xu et al. 1992) using a B1 field of 8.3 kHz for the TOCSY spin-lock as well as from a 70ms 13C-resolved NOESY experiment. The aromatic ring systems were correlated with β-carbons via the HBCBCGCDHD and HBCBCGCDCEHE experiments introduced by Kay (Yamazaki, Formankay et al. 1993). Assignment within the aromatic moieties was done using a 23 ms constant-time HMQC-TOCSY experiment (Zerbe, Szyperski et al. 1996), in which the proton TOCSY relay was tuned for direct (12ms) and relayed (40ms) transfer. Chemical shifts were finally picked in the [$^{15}N,^{1}H$]-HSQC and the 13.3ms constant-time [$^{13}C,^{1}H$]-HSQC spectra and indirectly referenced to the water

line at 4.63 ppm using the conversion factors of 0.10132900 (^{15}N) and 0.25144954 (^{13}C) (Cavanagh, Fairbrother et al. 1996). All experiments employed pulsed-field gradients.

The extent of amide hydrogen exchange was probed by recording a [^{15}N,^{1}H]-HSQC experiment, in which low-power irradiation was applied on the water resonance during the relaxation delay. Peak volumes were determined in this experiment and their values relative to a reference experiment conducted in the absence of irradiation but with otherwise identical parameter computed. ^{15}N-Relaxation data were recorded using proton detected version of the ^{15}N{^{1}H} steady-state NOE experiment(Noggle and Schirmer 1971). A recycle delay of 2.7s was used for the ^{15}N{^{1}H} NOE experiment and 128 scans were recorded per increment.

Data were usually extended by a factor of two using linear-prediction in the indirect dimensions and processed within the Bruker spectrometer software TOPSPIN 1.3. Integration of peak volumes was performed within the program SPSCAN. $^{3}J_{HN}$ scalar coupling constants were derived from the splitting of the in-phase doublets of [^{15}N,^{1}H]-HSQC peaks using the INFIT algorithm (Szyperski, Güntert et al. 1992) in XEASY (Bartels, Xia et al. 1995). Processed data were transferred into the program CARA for data analysis (Keller 2004).

Structure calculation

Distance restraints were obtained from NOESY spectra recorded with a mixing time of 70 ms either in 90% H_2O/10% $^{2}H_2O$ (^{15}N-resolved NOESY) or in 99.9% $^{2}H_2O$(^{13}C-resolved NOESY). In addition dihedral angle restraints were derived from TALOS(Cornilescu, Delaglio et al. 1999) using 13C chemical shifts of C_α and C_β atoms, and further such restraints were added by the CANDID/ATNOS suite of programs. Structures were calculated using a simulated-annealing protocol for molecular dynamics in torsion angle space as implemented in the program CYANA(Guntert, Mumenthaler et al. 1997; Guntert 2004). The final CYANA calculation was performed with 100 randomized starting structures, and the 20 CYANA conformers with the lowest target function values were selected to present the NMR ensemble. The conformers were

analyzed, including calculation of RMSD values, and figures were prepared within the program MOLMOL(Koradi, Billeter et al. 1996).

4.5. References:

Arshava, B., I. Taran, et al. (2002). "High resolution NMR analysis of the seven transmembrane domains of a heptahelical receptor in organic-aqueous medium." Biopolymers 64(3): 161-76.

Bader, R., A. Bettio, et al. (2001). "Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection." J. Mol. Biol. 305: 307-392.

Bader, R. and O. Zerbe (2005). "Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state?" Chembiochem 6(9): 1520-34.

Ballesteros, J. A., L. Shi, et al. (2001). "Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors." Mol. Pharmacol. 60(1): 1-19.

Bartels, C., T.-h. Xia, et al. (1995). "The program XEASY for computer-supported spectral analysis of biological macromolecules." J. Biomol. NMR 6: 1-10.

Bax, A., M. Clore, et al. (1990). "^{1}H-^{1}H Correlation via Isotropic Mixing of ^{13}C Magnetization, a New Three-Dimensional Approach for Assigning ^{1}H and ^{13}C Spectra of ^{13}C Enriched Proteins." J. Magn. Reson. 88: 425-431.

Bockmann, R. A. and A. Caflisch (2005). "Spontaneous formation of detergent micelles around the outer membrane protein OmpX." Biophys. J. 88(5): 3191-204.

Bright, J. N. and M. S. P. Sansom (2003). "The Flexing/Twirling Helix: Exploring the Flexibility about Moleular Hinges Formed by Proline and Glycine Motifs in Transmembrane Helices." J. Phys. Chem. 107: 627-636.

Brown, L. R. and K. Wuthrich (1981). "Melittin bound to dodecylphosphocholine micelles. H-NMR assignments and global conformational features." Biochim. Biophys. Acta 647(1): 95-111.

Bryson, K., L. J. McGuffin, et al. (2005). "Protein structure prediction servers at University College London." Nucleic Acids Res. 33(Web Server issue): W36-8.

Cavanagh, J., W. J. Fairbrother, et al. (1996). Protein NMR Spectroscopy: Principles and Practice. San Diego, Academic Press.

Chen, Q. and J. B. Konopka (1996). "Regulation of the G-protein-coupled alpha-factor pheromone receptor by phosphorylation." Mol. Cell Biol. 16(1): 247-57.

Clore, G. M. and A. M. Gronenborn (1991). "Applications of Three- and Four-Dimensional Heteronuclear NMR Spectroscopy to Protein Structure Determination." Prog. Nucl. Magn. Reson. Spectrosc. 23: 43.

Cornilescu, G., F. Delaglio, et al. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J. Biomol. NMR 13(3): 289-302.

Damberg, P., J. Jarvet, et al. (2001). "Micellar systems as solvents in peptide and protein structure determination." Methods Enzymol. 339: 271-85.

Dill, K. A. and P. J. Flory (1981). "Molecular organization in micelles and vesicles." Proc. Natl. Acad. Sci. U S A 78: 676-680.

Dyson, H. J., M. Rance, et al. (1988). "Folding of Immunogenic Peptide Fragments of Proteins in Water Solution." J. Mol. Biol. 201: 161-200.

Englander, J., L. Cohen, et al. (2006). "Selective labeling of a membrane peptide with (15)N-amino acids using cells grown in rich medium." Biopolymers.

Estephan, R., J. Englander, et al. (2005). "Biosynthesis and NMR analysis of a 73-residue domain of a Saccharomyces cerevisiae G protein-coupled receptor." Biochemistry 44(35): 11795-810.

Fernandez, C., C. Hilty, et al. (2002). "Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy." Proc. Natl. Acad. Sci. U S A 99(21): 13533-7.

Fesik, S. W. and E. R. P. Zuiderweg (1990). "Heteronuclear Three-Dimensional NMR Spectroscopy of Isotopically Labelled Biological Macromolecules." Q. Rev. Biophys. 23: 97.

Flower, D. R. (1999). "Modelling G-protein-coupled receptors for drug design." Biochim. Biophys. Acta 1422(3): 207-34.

Grisshammer, R., J. F. White, et al. (2005). "Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview." J. Struct. Funct. Genomics 6(2-3): 159-63.

Grzesiek, S. and A. Bax (1992). "Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR." J. Am. Chem. Soc. 114: 6291-6293.

Guntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol. Biol. 278: 353-78.

Guntert, P., C. Mumenthaler, et al. (1997). "Torsion Angle Dynamics For Nmr Structure Calculation With the New Program Dyana." Journal of Molecular Biology 273(1): 283-298.

Hessa, T., H. Kim, et al. (2005). "Recognition of transmembrane helices by the endoplasmic reticulum translocon." Nature 433(7024): 377-381.

Hilty, C., G. Wider, et al. (2004). "Membrane protein-lipid interactions in mixed micelles studied by NMR spectroscopy with the use of paramagnetic reagents." Chembiochem 5(4): 467-73.

Ji, T. H., M. Grossmann, et al. (1998). "G protein-coupled receptors. I. Diversity of receptor-ligand interactions." J. Biol. Chem. 273(28): 17299-302.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J. Mol. Biol. 292(2): 195-202.

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers 22(12): 2577-637.

Katragadda, M., J. L. Alderfer, et al. (2001). "Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin." Biophys. J. 81(2): 1029-36.

Keller, R. L. J. (2004). The Computer Aided Resoance Assignment Tutorial, Cantina Verlag.

Killian, J. A., T. P. Trouard, et al. (1994). "A general method for the preparation of mixed micelles of hydrophobic peptides and sodium dodecyl sulphate." FEBS Lett. 348(2): 161-5.

Koradi, R., M. Billeter, et al. (1996). "Molmol - a Program for Display and Analysis of Macromolecular Structures." J. Mol. Graph. 14(1): 51.

Krishna, A. G., S. T. Menon, et al. (2002). "Evidence that helix 8 of rhodopsin acts as a membrane-dependent conformational switch." Biochemistry 41(26): 8298-309.

Krueger-Koplin, R. D., P. L. Sorgen, et al. (2004). "An evaluation of detergents for NMR structural studies of membrane proteins." J. Biomol. NMR 28(1): 43-57.

Ladokhin, A. S. and S. H. White (1999). "Folding of amphipathic alpha-helices on membranes: energetics of helix formation by melittin." J. Mol. Biol. 285(4): 1363-9.

Lundstrom, K. (2005). "The future of G protein-coupled receptors as targets in drug discovery." IDrugs 8(11): 909-13.

MacKenzie, K. R., J. H. Prestegard, et al. (1997). "A transmembrane helix dimer: Structure and implications." Science 276: 131.

Melnyk, R. A., A. W. Partridge, et al. (2003). "Polar residue tagging of transmembrane peptides." Biopolymers 71(6): 675-85.

Mombaerts, P. (1999). "Seven-transmembrane proteins as odorant and chemosensory receptors." Science 286(5440): 707-11.

Montelione, G. T. (1992). "An efficient triple-resonance experiment using 13C isotropic mixing for determining sequence-specific assignments of isotopically enriched proteins." J. Am. Chem. Soc. 114: 10974-10975.

Naider, F., B. Arshava, et al. (2001). "Peptide fragments as models to study the structure of a G-protein coupled receptor: the alpha-factor receptor of Saccharomyces cerevisiae." Biopolymers 60(5): 334-50.

Naider, F. and J. M. Becker (2004). "The alpha-factor mating pheromone of Saccharomyces cerevisiae: a model for studying the interaction of peptide hormones and G protein-coupled receptors." Peptides 25(9): 1441-63.

Naider, F., F. X. Ding, et al. (2003). "Synthesis and biophysical characterization of a multidomain peptide from a Saccharomyces cerevisiae G protein-coupled receptor." J. Biol. Chem. 278(52): 52537-45.

Naider, F., S. Khare, et al. (2005). "Synthetic peptides as probes for conformational preferences of domains of membrane receptors." Biopolymers 80(2-3): 199-213.

Noggle, J. H. and R. E. Schirmer (1971). The Nuclear Overhauser Effect - Chemical Applications. New York, Academic Press.

Olejniczak, E. T., R. X. Xu, et al. (1992). "A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins." J. Biomol. NMR 2(6): 655-9.

Palczewski, K., T. Kumasaka, et al. (2000). "Crystal structure of rhodopsin: A G protein-coupled receptor." Science 289(5480): 739-45.

Papavoine, C. H., J. M. Aelen, et al. (1995). "NMR studies of the major coat protein of bacteriophage M13. Structural information of gVIIIp in dodecylphosphocholine micelles." Eur. J. Biochem. 232(2): 490-500.

Park, S. H., S. Prytulla, et al. (2006). "High-resolution NMR spectroscopy of a GPCR in aligned bicelles." J. Am. Chem. Soc. 128(23): 7402-3.

Pellegrini, M. and D. F. Mierke (1999). "Structural characterization of peptide hormone/receptor interactions by NMR spectroscopy." Biopolymers 51(3): 208-20.

Pellegrini, M., M. Royo, et al. (1996). "Conformational characterization of a peptide mimetic of the third cytoplasmic loop of the G-protein coupled parathyroid hormone/parathyroid hormone related protein receptor." Biopolymers 40(6): 653-66.

Pervushin, K. V., V. Orekhov, et al. (1994). "Three-dimensional structure of (1-71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by 15N-1H heteronuclear NMR spectroscopy." Eur. J. Biochem. 219(1-2): 571-83.

Popot, J. L. and D. M. Engelman (2000). "Helical membrane protein folding, stability, and evolution." Annu. Rev. Biochem. 69: 881-922.

Sanders, C. R., A. K. Hoffmann, et al. (2004). "French swimwear for membrane proteins." Chembiochem 5(4): 423-426.

Sarramegn, V., I. Muller, et al. (2006). "Recombinant G protein-coupled receptors from expression to renaturation: a challenge towards structure." Cell Mol. Life Sci. 63(10): 1149-64.

Sarramegna, V., F. Talmont, et al. (2003). "Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification." Cell Mol. Life Sci. 60(8): 1529-46.

Sass, H. J., G. Musco, et al. (2000). "Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes." J. Biomol. NMR 18(4): 303-9.

Strader, C. D., T. M. Fong, et al. (1994). "Structure and function of G protein-coupled receptors." Annu. Rev. Biochem. 63: 101-32.

Szyperski, T., P. Güntert, et al. (1992). "Determination of Scalar Coupling Constants by Inverse Fourier Transfomation of In-Phase Multiplets." J. Magn. Reson. 99: 552-560.

Thompson, M. D., W. M. Burnham, et al. (2005). "The G protein-coupled receptors: pharmacogenetics and disease." Crit. Rev. Clin. Lab. Sci. 42(4): 311-92.

Tian, C., R. M. Breyer, et al. (2005). "Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor." J. Am. Chem. Soc. 127(22): 8010-1.

Tian, C., R. M. Breyer, et al. (2006). "Solution NMR spectroscopy of the human vasopressin v2 receptor, a g protein-coupled receptor" J. Am. Chem. Soc. 128(15): 5300.

Tieleman, D. P., D. van der Spoel, et al. (2000). "Molecular Dynamics Simulation of Dodecylphosphocholine Micelles at Different Aggregate Sizes: Micellar Structure and Chain Relaxation." J. Phys. Chem. 104: 6380-6388.

Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science 278(5340): 1111-4.

Valentine, K. G., S. F. Liu, et al. (2001). "Structure and topology of a peptide segment of the 6th transmembrane domain of the Saccharomyces cerevisae alpha-factor receptor in phospholipid bilayers." Biopolymers 59(4): 243-56.

Vold, R. R., R. S. Prosser, et al. (1997). "Isotropic solutions of phospholipid bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides." J. Biomol. NMR 9(3): 329-35.

White, S. H. and W. C. Wimley (1998). "Hydrophobic interactions of peptides with membrane interfaces." Biochim. Biophys. Acta 1376(3): 339-52.

White, S. H. and W. C. Wimley (1999). "Membrane protein folding and stability: physical principles." Annu. Rev. Biophys. Biomol. Struct. 28: 319-65.

Wiener, M. C. and S. H. White (1992). "Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure." Biophys. J. 61(2): 437-47.

Wimley, W. C. and S. H. White (1996). "Experimentally determined hydrophobicity scale for proteins at membrane interfaces." Nature Struct. Biol. 3: 842-848.

Wittekind, M. and L. Mueller (1993). "HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins." J. Magn. Res. Series B 101(2): 201-205.

Xie, H., F. X. Ding, et al. (2000). "Synthesis and biophysical analysis of transmembrane domains of a Saccharomyces cerevisiae G protein-coupled receptor." Biochemistry 39(50): 15462-74.

Yamazaki, T., J. D. Formankay, et al. (1993). "2-Dimensional Nmr Experiments for Correlating C-13-Beta and H-1-Delta/Epsilon Chemical-Shifts of Aromatic Residues in C-13-Labeled Proteins Via Scalar Couplings." J. Am. Chem. Soc. 115(23): 11054-11055.

Yeagle, P. L. and A. D. Albert (2002). "Use of nuclear magnetic resonance to study the three-dimensional structure of rhodopsin." Methods Enzymol. 343: 223-31.

Yeagle, P. L., J. L. Alderfer, et al. (1997). "Three-Dimensional Structure Of the Cytoplasmic Face Of the G Protein Receptor Rhodopsin." Biochemistry 36(32): 9649-9654.

Yeagle, P. L., J. L. Alderfer, et al. (1997). "The First and Second Cytoplasmic Loops Of the G-Protein Receptor, Rhodopsin, Independently Form Beta-Turns." Biochemistry 36(13): 3864-3869.

Yeagle, P. L., G. Choi, et al. (2001). "Studies on the structure of the G-protein-coupled receptor rhodopsin including the putative G-protein binding site in unactivated and activated forms." Biochemistry 40(39): 11932-7.

Zerbe, O., T. Szyperski, et al. (1996). "Three-Dimensional H-1-TOCSY-Relayed ct-[C-13,H-1]-HMQC For Aromatic Spin System Identification In Uniformly C-13-Labeled Proteins." J. Biomol. NMR 7(2): 99-106.

Zhao, J., H. Zheng, et al. (2006). "NMR Characterization of Recombinant Transmembrane Protein CB2 Fragment CB2180-233." Protein Pept. Let. 13: 335-342.

Zheng, H., J. Zhao, et al. (2006). "A transmembrane helix-bundle from G-protein coupled receptor CB2: biosynthesis, purification, and NMR characterization." Biopolymers 83(1): 46-61.

4.6. Supplementary material

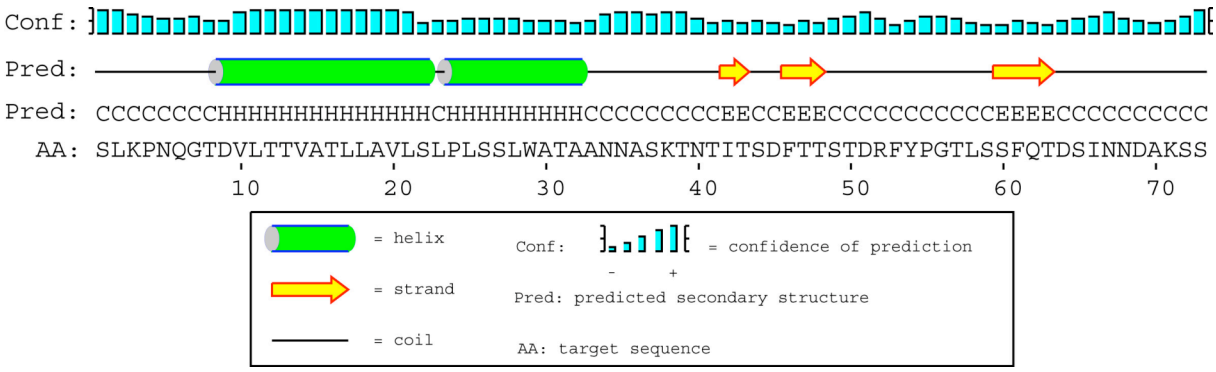


Fig. S1: Secondary structure prediction for EL3-TM7-CT40 based on position-specific scoring matrices

Table S1: Chemical shifts of ^{15}N and ^{1}H resonances for 0.5 mM EL3-TM7-CT40 in 90% H_2O/10%$^{2}H_2O$, pH 6.0, T 310K, referenced to the signal of residual HDO at 4.63 ppm, and ^{13}C and ^{15}N nuclei are indirectly referenced to that (see Materials & Methods).

	N	H^N	H^{α}	H^{β}	Others
Ser 1	-	-	-	-	γOH -
Leu 2	-	-	-	-	γH -; δCH_3 -
Lys 3	122.16	8.30	4.59	1.69, 1.84	γCH_2 1.48; δCH_2 1.70; εCH_2 3.00; ζNH_3^+ -
Pro 4	-		4.41	1.94, 2.28	γCH_2 2.04; δCH_2 3.68, 3.81
Asn 5	117.63	8.58	4.66	2.84	δNH_2 -
Gln 6	119.47	8.30	4.37	2.05, 2.16	γCH_2 2.37; εNH_2 -
Gly 7	109.21	8.56	3.99		
Thr 8	114.00	8.10	4.23	4.30	γCH_3 1.21; γOH -
Asp 9	122.86	8.43	4.70	2.83	δH -
Val 10	122.19	8.59	3.78	2.21	γCH_3 0.97, 1.09
Leu 11	119.91	8.36	4.05	1.89	γH 1.66; δCH_3 0.90
Thr 12	114.88	8.13	3.98	4.24	γCH_3 1.26; γOH -

Thr 13	120.27	7.96	3.95	4.37	γCH_3 1.22; γOH -
Val 14	120.32	8.42	3.58	2.24	γCH_3 0.96, 1.07
Ala 15	120.81	8.37	4.00	1.57	
Thr 16	114.05	8.05	3.97	4.40	γCH_3 1.27; γOH -
Leu 17	121.51	7.93	4.12	1.87	γH 1.69; δCH_3 0.90
Leu 18	116.52	8.33	4.01	1.88	γH 1.59; δCH_3 0.88
Ala 19	119.90	7.79	4.16	1.58	
Val 20	116.00	7.83	4.00	2.35	γCH_3 1.03, 1.14
Leu 21	117.61	7.84	4.21	1.88	γH -; δCH_3 0.89
Ser 22	111.97	7.84	4.13	3.98	γOH -
Leu 23*	121.3	7.97	4.15	1.88	γH 1.58; δCH_3 0.90
Pro 24	-		4.41	2.28	γCH_2 1.99; δCH_2 3.66, 3.80
Leu 25*	121.09	8.23	4.60	1.58	γH 1.27; δCH_3 0.88
Ser 26*	114.39	8.12	4.37	4.23	γOH -
Ser 27	117.12	7.94	4.41	3.99	γOH -
Leu 28	122.20	7.94	4.15	1.88	γH 1.58; δCH_3 0.89
Trp 29	117.68	8.06	4.45	3.36, 3.45	δ^1H 7.25; ε^1NH 10.51; ε^3H 7.49; ζ^2H 7.47; ζ^3H 6.97; η^2H 7.08
Ala 30	120.34	7.91	4.18	1.51	
Thr 31	112.90	7.95	4.24	4.16	γCH_3 1.21; γOH -
Ala 32	124.78	8.15	4.12	1.35	
Ala 33	120.00	8.25	4.05	1.21	
Asn 34	115.79	8.04	4.62	2.84	δNH_2 -
Asn 35	118.00	8.08	4.65	2.73, 2.82	δNH_2 -
Ala 36	122.96	8.10	4.26	1.43	
Ser 37	113.13	8.13	4.40	3.94	γOH -
Lys 38	121.83	8.15	4.39	1.84, 1.93	γCH_2 1.47; δCH_2 1.70; εCH_2 3.00; ζNH_3^+ -

Thr 39	112.98	8.02	4.28	4.24	γCH_3 1.21; γOH -
Asn 40	120.28	8.36	4.77	2.84	δNH_2 -
Thr 41	113.71	8.10	4.34	4.24	γCH_3 1.20; γOH -
Ile 42	121.77	8.20	4.21	1.95	γCH_2 1.53; γCH_3 1.22; δCH_3 0.93
Thr 43	114.91	8.08	4.31	-	γCH_3 1.20; γOH -
Ser 44	116.80	8.14	4.39	3.89	γOH -
Asp 45	121.37	8.25	4.60	2.62	δH -
Phe 46	119.56	8.19	4.58	3.12, 3.20	δH 7.28; εH 7.29; ζH 7.22
Thr 47	113.17	8.06	4.29	-	γCH_3 1.21; γOH -
Thr 48	114.42	8.02	4.36	4.29	γCH_3 1.21; γOH -
Ser 49	116.94	8.17	4.51	3.90	γOH -
Thr 50	-	-	-	-	γCH_3 -; γOH -
Asp 51	121.28	8.20	4.48	2.65	δH -
Arg 52	118.18	7.86	4.19	1.62	γCH_2 1.38; δCH_2 3.07; εNH -; ηNH_2 -
Phe 53	118.03	8.00	4.58	2.98, 3.12	δH 7.24; εH 7.24; ζH 7.20
Tyr 54	120.04	7.87	4.72	2.88, 3.02	δH 7.11; εH 6.82; ηOH -
Pro 55	-		4.42	2.28	γCH_2 1.95; δCH_2 3.40, 3.67
Gly 56	107.76	8.08	3.98		
Thr 57	113.55	8.00	4.29	4.20	γCH_3 1.21; γOH -
Leu 58	122.14	8.32	4.40	1.62	γH -; δCH_3 0.91
Ser 59	114.48	8.15	4.38	3.89	γOH -
Ser 60	116.22	8.10	4.37	3.78	γOH -
Phe 61	120.18	8.00	4.59	3.06, 3.20	δH 7.30; εH 7.30; ζH 7.23
Gln 62	119.98	8.04	4.43	2.02, 2.15	γCH_2 2.33; εNH_2 -
Thr 63	114.5	8.19	4.30	4.25	γCH_3 1.21; γOH -

Asp 64	121.99	8.37	4.65	2.65	δH -
Ser 65	115.32	8.20	4.42	3.85	γOH -
Ile 66	121.23	8.02	4.16	1.89	γCH_2 1.46; γCH_3 1.17; δCH_3 0.89
Asn 67	121.28	8.36	4.71	2.71, 2.84	δNH_2 -
Asn 68	119.65	8.36	4.69	2.73, 2.84	δNH_2 -
Asp 69	120.32	8.29	4.57	2.64, 2.70	δH -
Ala 70	123.53	8.08	4.30	1.39	
Lys 71	119.87	8.18	4.39	1.80, 1.89	γCH_2 1.46; δCH_2 1.71; εCH_2 3.00; ζNH_3^+ -
Ser 72	116.95	8.29	4.51	3.92	γOH -
Ser 73	122.57	7.93	-	3.85	γOH -

* the assignments from these residues were done based on the NOESY data, because a combination of overlap and missing peaks prevented unambiguous assignments based on the triple-resonance data.

Table S2: ^{13}C chemical shift table

	N	C^α	C^β	Others
Ser 1	-	-	-	
Leu 2	-	-	-	γCH -; δCH_3 -
Lys 3	122.16	53.88	32.71	γCH_2 24.81; δCH_2 28.93; εCH_2 41.88
Pro 4	-	63.41	32.00	γCH_2 27.49; δCH_2 50.32
Asn 5	117.63	53.47	38.65	γC -
Gln 6	119.47	55.96	29.70	γCH_2 33.92; δC -
Gly 7	109.21	45.57		
Thr 8	114.00	62.91	69.06	γCH_3 22.06
Asp 9	122.86	54.37	41.04	γC -
Val 10	122.19	65.77	31.74	γCH_3 21.55, 22.68

Leu 11	119.91	58.04	41.05	γCH 27.05; δCH_3 24.01, 24.06
Thr 12	114.88	66.16	67.93	γCH_3 21.80
Thr 13	120.27	67.22	68.05	γCH_3 22.02
Val 14	120.32	66.87	31.45	γCH_3 21.72, 23.30
Ala 15	120.81	55.44	18.42	
Thr 16	114.05	66.22	68.40	γCH_3 21.79
Leu 17	121.51	57.53	41.82	γCH 27.02; δCH_3 23.96
Leu 18	116.52	57.40	41.25	γCH -; δCH_3 23.25
Ala 19	119.90	54.64	18.41	
Val 20	116.00	64.47	31.84	γCH_3 21.43, 22.09
Leu 21	117.61	55.54	42.02	γCH -; δCH_3 25.86
Ser 22	111.97	64.20	62.87	
Leu 23	121.3	56.83	41.77	γCH 18.46; δCH_3 23.90
Pro 24	-	63.18	31.96	γCH_2 27.55; δCH_2 50.21
Leu 25	121.09	53.89	41.25	γCH -; δCH_3 -
Ser 26	114.39	61.97	63.19	
Ser 27	117.12	60.73	63.12	
Leu 28	122.20	54.65	41.58	γCH -; δCH_3 25.25
Trp 29	117.68	58.99	29.54	γC -; δ^1CH 127.10; δ^2C -; ε^2C -; ε^3CH 120.49; ζ^2CH 114.40; ζ^3CH 121.11; η^2CH 123.75
Ala 30	120.34	54.02	18.81	
Thr 31	112.90	62.66	69.19	γCH_3 21.91
Ala 32	124.78	53.73	18.90	
Ala 33	120.00	53.33	18.58	
Asn 34	115.79	53.78	38.80	γC -
Asn 35	118.00	54.14	39.35	γC -
Ala 36	122.96	53.23	19.07	
Ser 37	113.13	58.99	63.64	

Lys 38	121.83	56.23	32.78	γCH_2 25.12; δCH_2 28.93; εCH_2 41.93
Thr 39	112.98	62.40	69.35	γCH_3 21.90
Asn 40	120.28	53.29	38.75	γC -
Thr 41	113.71	61.89	69.37	γCH_3 21.90
Ile 42	121.77	61.50	38.43	γCH_2 27.78; γCH_3 -; δCH_3 17.80
Thr 43	114.91	61.94	69.42	γCH_3 21.94
Ser 44	116.80	59.20	63.81	
Asp 45	121.37	54.31	40.61	γC -
Phe 46	119.56	58.10	39.49	γC -; δCH 132.06; εCH 131.37; ζCH 129.32
Thr 47	113.17	62.54	69.50	γCH_3 21.91
Thr 48	114.42	62.09	69.15	γCH_3 22.1
Ser 49	116.94	58.73	63.60	
Thr 50	-	-	-	γCH_3
Asp 51	121.28	54.82	40.65	γC -
Arg 52	118.18	56.05	30.58	γCH_2 26.82; δCH_2 43.19; ζC -
Phe 53	118.03	57.46	39.79	γC -; δCH 131.94; εCH 131.48; ζCH 129.32
Tyr 54	120.04	55.73	38.56	γC -; δCH 133.20; εCH 118.07; ζC -
Pro 55	-	63.41	32.12	γCH_2 27.35; δCH_2 50.40
Gly 56	107.76	45.63		
Thr 57	113.55	62.81	69.67	γCH_3 21.86
Leu 58	122.14	55.12	42.41	γCH -; δCH_3 25.59
Ser 59	114.48	59.30	63.31	
Ser 60	116.22	58.99	63.22	
Phe 61	120.18	57.90	39.41	γC -; δCH 132.15; εCH 131.29; ζCH 129.40

Gln 62	119.98	55.50	29.77	γCH_2 33.90; δC -
Thr 63	114.5	62.21	69.22	γCH_3 21.86-
Asp 64	121.99	54.11	40.88	γC -
Ser 65	115.32	58.40	63.47	
Ile 66	121.23	61.09	38.64	γCH_2 27.43; γCH_3 -; δCH_3 17.58
Asn 67	121.28	52.84	38.87	γC -
Asn 68	119.65	53.24	39.14	γC -
Asp 69	120.32	54.22	40.68	γC -
Ala 70	123.53	52.52	18.95	
Lys 71	119.87	56.31	32.93	γCH_2 24.78; δCH_2 28.9; εCH_2 41.87
Ser 72	116.95	58.53	63.44	
Ser 73	122.57	59.69	64.42	

5. Structure of a double transmembrane fragment of a G-protein coupled receptor in micelles

The structure and dynamic properties of an eighty-residue fragment of Ste2p, the G-protein coupled receptor for α-factor of *Saccharomyces cerevisiae,* was studied in lyso-palmitoylphosphatidylglycerol (LPPG) micelles using solution NMR spectroscopy. The fragment Ste2p(31-110) (TM1-TM2) consisted of 19 residues from the N-terminal domain, the 1st transmembrane helix (TM1), the first cytoplasmic loop, the second transmembrane helix (TM2) and 7 residues from the first extracellular loop. Multidimensional NMR experiments on [^{15}N], [^{15}N, ^{13}C], [^{15}N, ^{13}C, ^{2}H]-labeled TM1-TM2 and on peptides selectively labeled at specific amino acid residues or protonated at selected methyl groups resulted in >95 % assignment of backbone and side chain nuclei. The NMR investigation revealed the secondary structure of specific residues of TM1-TM2. TALOS constraints and NOE connectivities were used to calculate a structure for TM1-TM2 that was highlighted by the presence of three α-helices encompassing residues 39-47, 49-72 and 80-103, with higher flexibility around the internal Arg58 site of TM1. RMSD values of individually superimposed helical segments 39-47, 49-72 and 80-103 were 0.25±0.10 Å, 0.40±0.13 Å and 0.57±0.19 Å, respectively. Several long-range interhelical connectivities supported the folding of TM1-TM2 into a tertiary structure typified by a crossed helix that splays apart toward the extracellular regions and contains considerable flexibility in the G^{56}VRSG^{60} region. ^{15}N-relaxation and hydrogen-deuterium exchange data support a stable fold for the transmembrane parts of TM1-TM2, whereas the solvent-exposed segments are more flexible. The NMR structure is consistent with the results of biochemical experiments that identified the ligand binding site within this region of the receptor.

submitted for publication: Neumoin, A., Cohen, L., Arshava, B., Tantry, S., Becker, J., Zerbe, O., Naider, F.

5.1. Introduction

Relatively few high resolution structures for membrane receptors and transporters have appeared in the protein database despite the fact that these integral membrane proteins (IMPs) have been estimated to constitute 25 to 30 % of eukaryotic proteins(1, 2). Among IMPs G protein coupled receptors (GPCRs) represent a biomedically important superfamily of eukaryotic proteins. Eight hundred GPCRs have been identified by analysis of the human genome and thirty to fifty percent of prescription drugs target these proteins (3). However, high-resolution structures reported to date are limited to those of rhodopsin (4) and the β-adrenergic receptors(5, 6). To alleviate this underrepresentation many laboratories are actively involved in developing new techniques to express, isolate and crystallize membrane proteins(7, 8), and to study these in membrane-like environments(9, 10). Crystallization of membrane transporters and receptors has been aided by construction of chimeras, the use of antibodies and mutation to decrease the inherent flexibility of these IMPs. It also has been greatly accelerated by the use of high throughput robot assisted methodologies. Although significant successes are being reported, crystallization of IMPs is still more art than science. Noteworthy progress has also been achieved in solid-state NMR studies (11-14), and most recently solution-state NMR experiments in detergent resulted in the complete backbone assignments of sensory rhodopsin II from *Natronomonas pharaonis* (15). Until crystallization of IMPs becomes more routine or there is a major breakthrough in solution/solid state NMR approaches, progress in studies of full-length transporters and receptors will likely be slow.

The use of peptides that represent individual regions of integral membrane proteins as surrogates for the partial structure of IMPs was predicated upon a model of membrane protein assembly(16, 17). In the first stage of the two-state model of IMP folding originally proposed by Engelman (16), upon partitioning into the membrane interface the peptide forms α-helical transmembrane domains (TMs), which then spontaneously insert into the bilayer core due to entropic driving forces. In stage two these independent domains assemble into the three-dimensional protein structure. TM-TM assembly likely involves van der Waals packing forces, a few polar or electrostatic interactions, and C-H---O=C hydrogen bonds, and often is influenced by GXXXG motifs

and the presence of proline residues(18). Later, the two-stage model was extended to include additional interactions with membrane-lipid head groups (17). Peptides corresponding to single transmembrane domains of bacteriorhodopsin(19-24), rhodopsin(25-29), Ste2p, the α-factor receptor(27, 30-35) and the adenosine A2 receptor(36, 37) were shown to assume helices in membrane mimetic solvents, thereby providing evidence for the two stage model.

Despite the extensive use of peptide fragments to understand biophysical properties of regions of IMPs significant skepticism remains concerning the biological significance of the information obtained from such investigations. Few studies have been conducted on peptides longer than a single TM. CD studies revealed that two TM fragments of the μ-opioid receptor and the CB2 cannabinoid receptor were highly helical in membrane mimetic solvents such as trifluoroethanol/water and a variety of detergents (38-40). Well-resolved 2D NMR spectra were measured and 80% of the peaks were assigned for the CB2 double TM fragment in DMSO solution. An NMR structure was reported for a two TM fragment of the human glycine receptor in trifluoroethanol (41), and a series of biophysical studies from the CFTR protein provided insights into the influence of turn structures and residue effects on helical hairpin formation(42). Nevertheless, no high-resolution information is presently available for a multitopic GPCR fragment in a lipid-like environment, and it is not clear whether two contiguous domains of these heptahelical receptors will pack to a stable tertiary structure in a detergent micelle in the absence of interactions with the remainder of the protein.

Herein we present a detailed high-resolution NMR study on an 80 residue fragment of the yeast α-factor receptor Ste2p(G31-T110) containing a short stretch of the N-terminus(NT)-transmembrane domain 1 (TM1), the first intracellular loop (IL1), transmembrane domain 2 (TM2), and a short stretch of the first extracellular loop (EL1) (Fig. 1). This polypeptide was biosynthesized with [^{15}N], [^{15}N,^{13}C], [^{15}N,^{13}C,^{2}H] uniform isotope labeling. In addition, it was labeled at specific amino acids, or at unique methyl protons in an otherwise perdeuterated background.

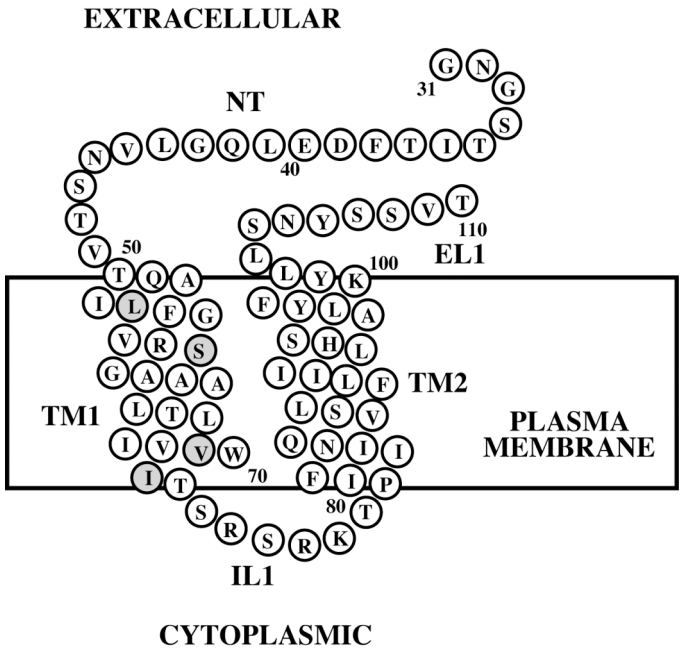


Figure 1: Schematic representation of Ste2p(G31-T110) [TM1-TM2]: NT: twenty residues of the N-terminal domain of Ste2p; TM1: first transmembrane helix; IL1: first intracellular loop; TM2: second transmembrane helix; EL1: first extracellular loop. The mutated methionines and cysteine are shaded. The numbering used follows that of the intact receptor.

Using triple-resonance NMR experiments nearly complete assignment of backbone and side chain resonances in 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG) micelles was accomplished. 3D NOESY spectra in combination with deuterated LPPG allowed assignment of a large number of medium-range NOEs that unambiguously established the secondary structure. Moreover, the use of a labeling pattern introduced by Kay and coworkers (43) allowed determination of several interhelical long-range NOEs between methyl groups, and these NOE-derived restraints were used to calculate a model of the structure of the 80-residue fragment. The structure represents the first high-resolution structure of a double transmembrane domain fragment of a GPCR in lipid. The data help to explain biochemical crosslinking studies that revealed an interaction of the 13th residue of the α-factor tridecapeptide with residues 58 and 59 of the α-factor receptor.

5.2. Results

Biosynthesis of Selectively Methyl Protonated Samples

Expression of selectively methyl protonated TM1-TM2 fusion protein in an otherwise deuterated background as described by Kay and coworkers (43) was performed using ketobutyric and ketoisovaleric acids that were isotopically labeled with ^{1}H, ^{2}H and ^{13}C as described by Tugarinov *et al.* (47). After the fusion protein was expressed, the cells were harvested, inclusion bodies were solubilized in 70% trifluoroacetic acid, and CNBr was used to remove the TrpDLE peptide segment immediately followed by purification by RP-HPLC in an acetonitrile:isopropanol:water gradient to >95% purity (Fig. 2) (44). The yield of purified peptide after lyophilization was 5.5 mg/L and the incorporation of ^{15}N, ^{13}C, and ^{2}H was greater than 95% (calculated MW=9702.29, observed MW=9660.42).

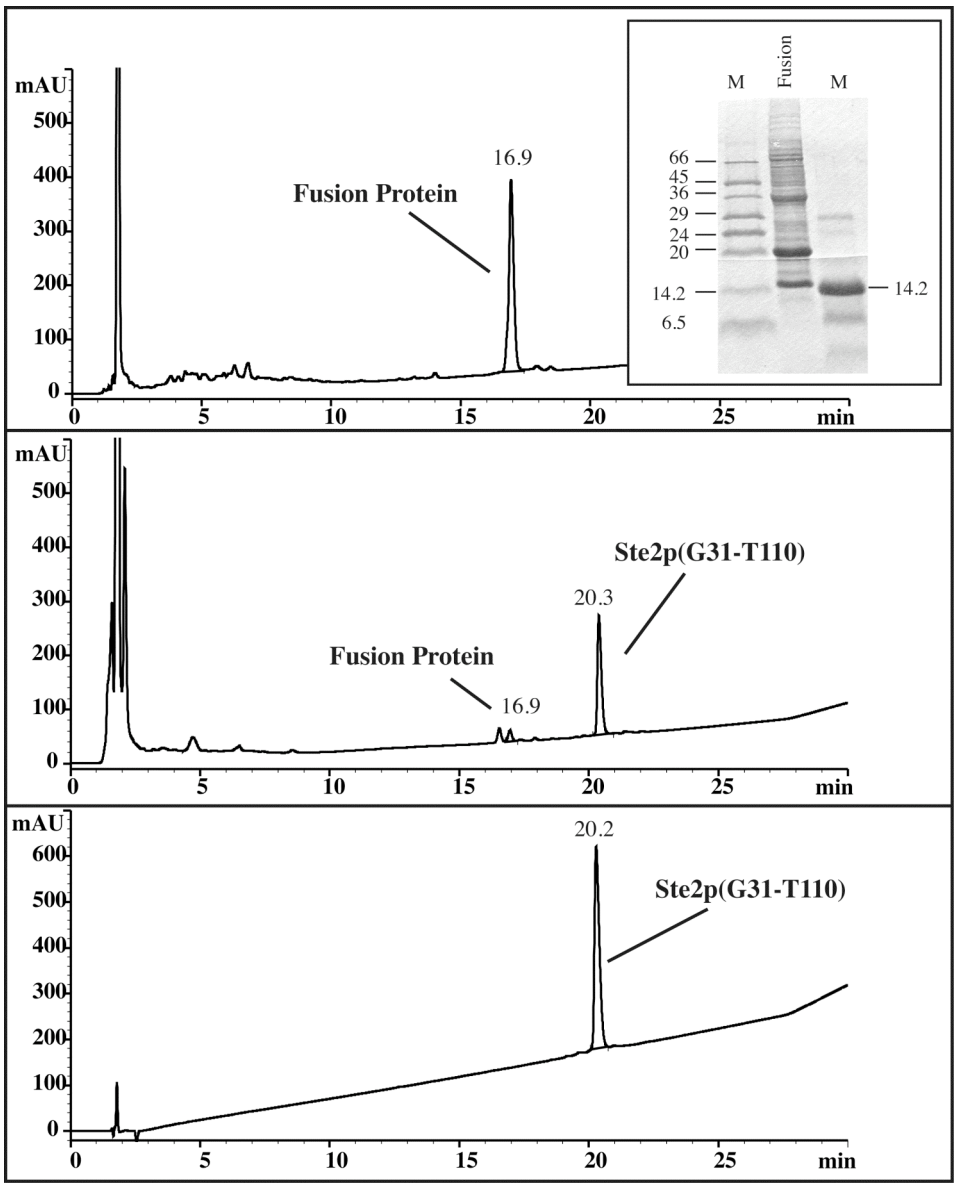


Figure 2: Cleavage and purification of the selectively methyl labelled Ste2p(G31-T110) peptide. Top panel: Inclusion bodies containing the fusion protein prior to cleavage with CNBr. Inset: SDS-PAGE gel of the inclusion bodies to show protein expression levels. Middle panel: Chromatogram of CNBr cleavage reaction after 1 hour. Lower panel: Analytical RP-HPLC of the purified Ste2p(G31-T110) after CNBr cleavage and purification on a preparative scale. Analytical reversed phase HPLC was performed with a 36-90% acetonitrile:water gradient with 10% isopropanol, 0.1% trifluoroacetic acid at 60°C on a Zorbax 300SB-C3 column.

Backbone resonance assignment

Resonance assignment of TM1-TM2 was accomplished using 3D triple-resonance NMR experiments. Seventy-five cross-peaks could be detected in the [^{15}N, ^{1}H]-HSQC spectrum (Fig. 3), for which well-separated resonances both in HNCO and HNCA spectra were present. Best results were obtained from the HNCO, HNCA, HN(CO)CA and HN(CA)CO experiments as the most sensitive triple-resonance experiments. About 30 of the crosspeaks in the [^{15}N, ^{1}H]-HSQC spectrum had corresponding intra- and interresidual Cα and Cβ peaks in the [^{13}C, ^{1}H] – strips of the HNCACB and CBCA(CO)NH experiments.

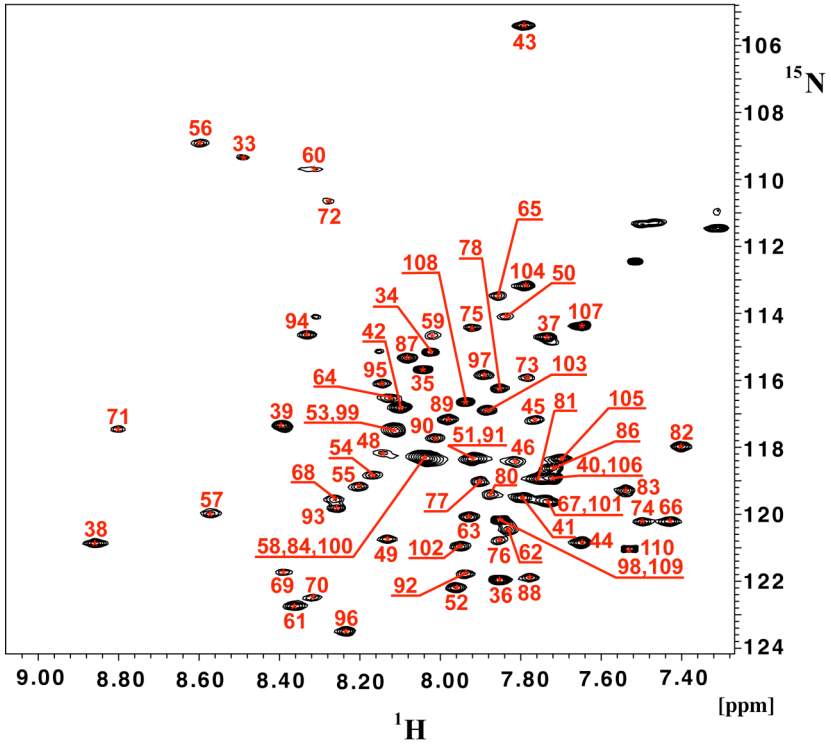


Figure 3: [^{15}N, ^{1}H]-HSQC spectrum of Ste2p(G31-T110). The spectrum was recorded at 700 MHz, 320K, on a 0.4mM sample of the peptide in 200 mM LPPG at pH 6.4. The sequence-specific assignments are annotated.

For about 20 peaks in the [^{15}N, ^{1}H]-HSQC spectrum one or more resonances were missing in the corresponding strips from HNCACB or CBCA(CO)NH, and for the remaining ~20 residues no peaks were observed in the corresponding strips. By searching

through [^{13}C, ^{1}H]–strips in the HNCA, HN(CO)CA and HNCO, HN(CA)CO spectra neighboring residues were identified (see Fig. S1). Data from the HNCACB and CBCA(CO)NH spectrum were used to assign and confirm the sequential assignments wherever possible. [^{15}N, ^{1}H]-HSQC spectra measured on TM1-TM2 selectively labeled with ^{15}N-amino-acids (Ile, Leu, Phe, Ala or Val) supported the assignments. Because of the large number of Ile and Leu residues in the TM1-TM2 sequence, particularly in the transmembrane domains, we experienced significant difficulties with the backbone assignment for residues 63 to 68 and 87 to 94. These difficulties could be resolved by searching for sequential HN-HN, HN-H(aliphatic) and H(aliphatic)-H(aliphatic) crosspeaks in the ^{15}N- and ^{13}C-resolved NOESY spectra. Subsequently these assignments were further cross-validated against hCCH-TOCSY and ^{13}C-resolved NOESY data. Because of overlapping or missing peaks in the triple-resonance spectra we were not able to assign amide moieties of Gly31, Asn32, Ser47 and Gln85. However, in combination with knowledge of typical chemical shifts encountered for such residues they could be assigned using the combination of the ct-[^{13}C, ^{1}H]-HSQC and ^{13}C-resolved NOESY. The overall completeness of the backbone assignment for Hα, Cα, HN and N was above 95%.

Side-chain resonance assignment

Based on the unique Hα and Cα resonance assignments derived from the backbone assignment process described above, the Hα/Cα cross-peaks were used as anchoring points for assignment of the aliphatic side chains. Because signal-to-noise in the hCCH-TOCSY spectrum was insufficient, extensive usage of ^{13}C-resolved NOESY and ^{15}N-resolved NOESY spectra was required. Due to extensive peak overlap we experienced severe difficulties for sidechain assignments of Ile and Leu residues, especially in the methyl region that was partially covered by peaks from the non-deuterated LPPG. However, when using partially deuterated d_{36}-LPPG signals cleared up and NOEs could be used for assignment purposes (Fig. S2). Moreover, we observed significant line narrowing of the crosspeaks most likely due to the reduced intermolecular dipolar broadening in the presence of the deuterated palmitoyl chain. To further facilitate

assignments a protein sample was prepared following methodology developed by Kay and coworkers (43), in which protonated methyl groups from Ile, Leu and Val are introduced into an otherwise completely perdeuterated background. Assignment of methyl groups of Val(Hγ), Ile(Hδ1) and Leu(Hδ1/δ2) residues was accomplished by using a HMCMCBCANH experiment (43) that correlates methyl group resonances of Val, Ile and Leu with the backbone amides. Even using this strategy the very narrow chemical shift dispersion of the methyl resonances of Leu residues in the ^{1}H-dimension (Fig. 4) resulted in ambiguous assignments of several methyl groups in the 700 MHz ^{13}C-resolved NOESY spectra such as for leucine residues 54, 64 and 97. Fortunately, resolution in the 900 MHz ^{13}C-resolved NOESY spectrum recorded on the selectively [^{15}N,^{13}C,^{2}H(^{1}H(methyl) - Ile, Leu, Val)] - labeled sample was sufficient to unambiguously establish a number of methyl-methyl NOEs that were critically important for establishing interhelical contacts and hence for orienting the two TM helices with the respect to each other.

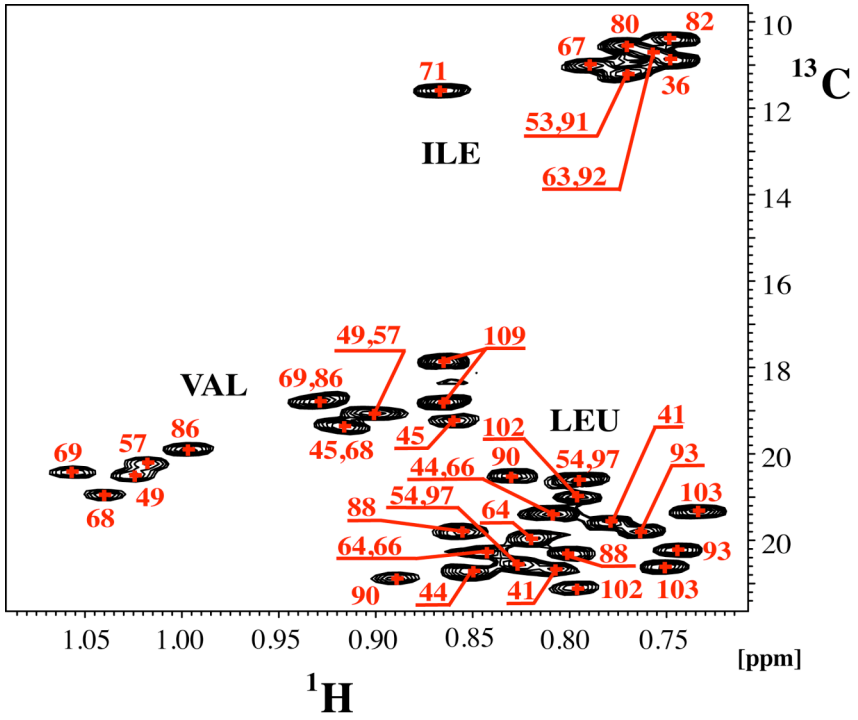


Figure 4: Methyl region from the 900 MHz ct-[^{13}C,^{1}H]-HSQC of the selectively methyl labelled Ste2p(G31-T110) sample in the d_{36}-LPPG solution. Assignments of methyl groups corresponding to Ile, Leu and Val residues are annotated.

The unique aromatic side-chain resonances of Trp70, His94, Tyr98, Tyr101 and Tyr106 were picked in the aromatic ct-[^{13}C, ^{1}H]-HSQC and assigned from 2D (HB)CB(CGCD)HD experiments. Aromatic protons from Phe residues 38, 55, 81, 89 and 99 were assigned using the ^{13}C-resolved aromatic NOESY experiment, but signal dispersion was so small that unambiguous assignments could only be established for Phe38. Generally no assignments were possible for amide and amino sidechain resonances. Ultimately, the percentage of the unambiguously assigned sidechain resonances was about 95%, while assignments of approximately 5% of the resonances including aromatic spin systems of phenylalanine residues 55, 81, 89 and 99 and aliphatic spin systems of leucine residues 54, 64 and 97 were ambiguous.

Backbone dynamics of TM1-TM2 in LPPG

Values of the heteronuclear ^{15}N(1H)-NOE (H-NOE) have been used to discriminate well-structured regions from those that display increased flexibility (56). High H-NOE values (> 0.6) are usually observed in elements of secondary structure as well as in rather rigid short loops. In the 2-TM protein subject to this study most residues in the segments 37-44, 50-72 and 80-101 had H-NOE values >0.75 (Fig. 5). The H-NOE values for many residues in the segments 45-49, 58-62 and 74-81 were reduced to around 0.6 indicating larger degrees of flexibility. The predicted transmembrane helices TM1 and TM2 are characterized by comparably large H-NOEs, while most values for residues S72 to P79, which by hydropathicity analysis are predicted to constitute the first intracellular loop, are lower. From residue Ser107 to the C-terminus and from Ile36 to the N-terminus the H-NOE values steadily decreased indicating that both termini are rather flexible. This conclusion is additionally supported by the absence of medium-range NOEs in these regions (*vide infra*).

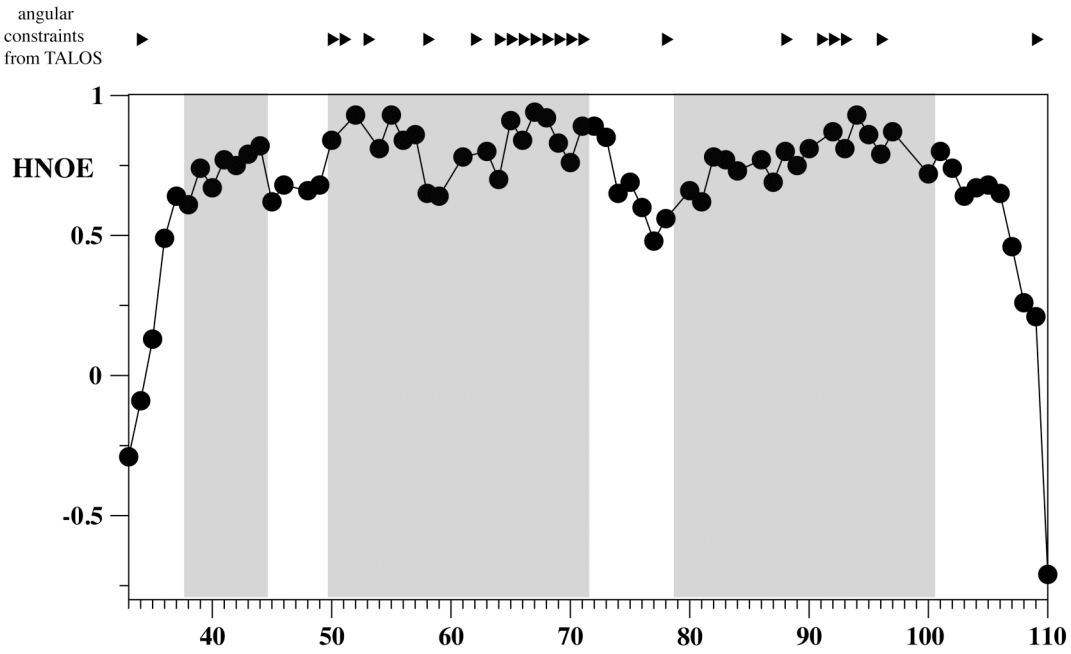


Figure 5: Values of the $^{15}N(^{1}H)$-NOE of Ste2p(G31-T110). The $^{15}N\{^{1}H\}$-NOEs were determined using spectra recorded at 700 MHz proton frequency. Highlighted with grey are the predicted α-helical regions of TM1-TM2. Residues, for which dihedral angle restraints calculated with TALOS were applied during the structure calculation, are marked by a filled triangle on top.

Structure calculation

Once the chemical shifts were available structure calculations were performed using the program CYANA and the internally implemented algorithm for automatic NOE assignment (61). The latter annotates peaks based on the match of chemical shifts, can apply multiple assignments in the form of ambiguous restraints, and uses network anchoring to select for correct long-range restraints. Distance restraints were obtained from ^{15}N-resolved NOESY and from ^{13}C-resolved NOESY spectra. In addition 42 dihedral angle restraints were derived from ^{13}C chemical shifts of Cα, Cβ,.C' and ^{15}N atoms using the program TALOS(60) (Fig. 5). Initially, hydrogen bond restraints were applied in the regions of the putative helices to facilitate automatic assignment of medium-range NOEs, in particular for automatically assigning i,i+3 contacts. After the

hydrogen bonds restraints were removed from the calculations, almost all medium-range contacts remained and these were manually checked in the 3D spectra to remove erroneous peak assignments resulting from the artificial restraints. As shown in Fig. 6 almost complete sets of α,β(i,i+3) and numerous α,N(i,i+3) and several α,N(i,i+4) contacts for the segments comprising residues 49-72 and 80-103, that correspond to the TM regions (TM1 50-72, TM2 79-103), were observed throughout the helices. Moreover, we observed characteristic α,β(i,i+3) contacts for the residues 39-47 of the N-terminal domain. This fragment of the peptide also appeared to be highly helical, forming an α-helix that is most probably surface associated (*vide infra*). When repeating the structure calculation in the absence of the artificial H-bond restraints an almost identical number of medium-range contacts were assigned in a highly similar fashion ruling out that the H-bond restraints were solely responsible for locating the TM helices.

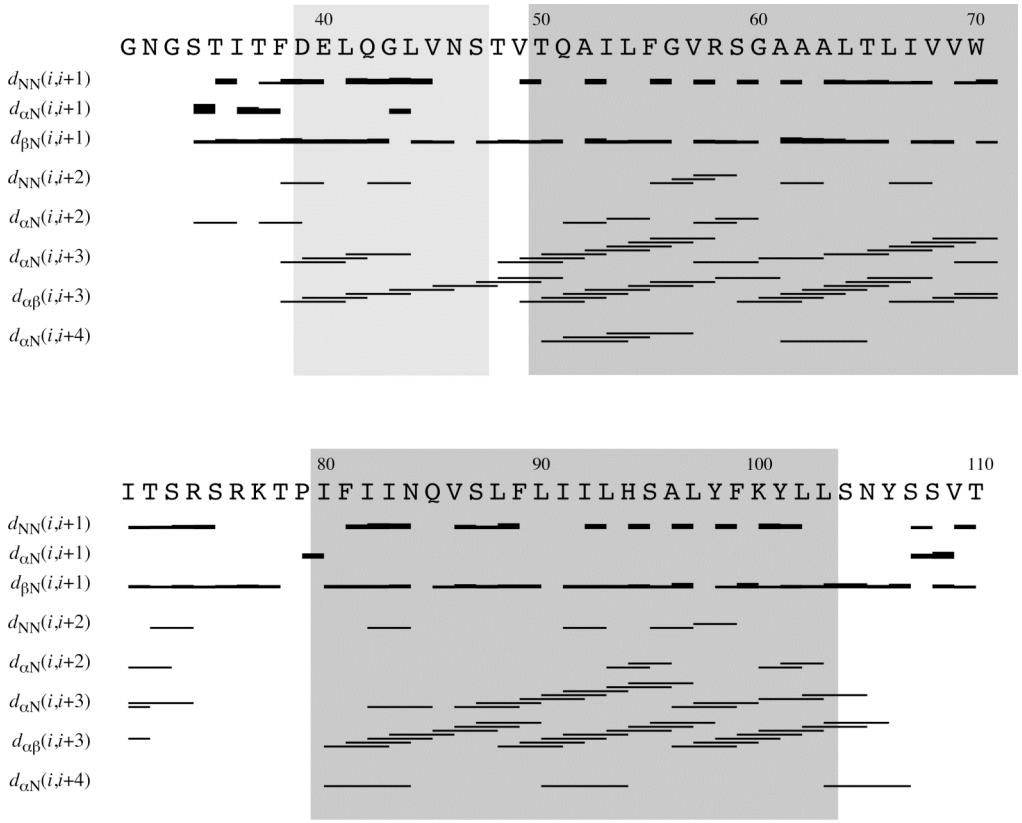


Figure 6: Sequence plot displaying characteristic upper distance restraints along the Ste2p(G31-T110) sequence derived from NOEs. Regions of the predicted TM helices and the extracellular helix are shaded in gray.

Using the above restraints (Fig. 6) and several long-range restraints measured on selectively methyl protonated samples we calculated the ensemble of the 20 lowest energy NMR conformations depicted in Fig. 7. The superposition of NMR-derived conformers reveals the presence of two TM helices corresponding to the predicted TM1 and TM2 segments, preceded by an additional helical region encompassing residues 39-47. In 11 out of the 20 conformers the first α-helix starts at residue 38 according to the criteria of Kabsch and Sanders (62). The TM1 and TM2 α-helices extend from residues 49-72 and 80-103 in all conformers, and in 9 out of 20 conformers the helices extend up to residues 73 and 106, respectively.

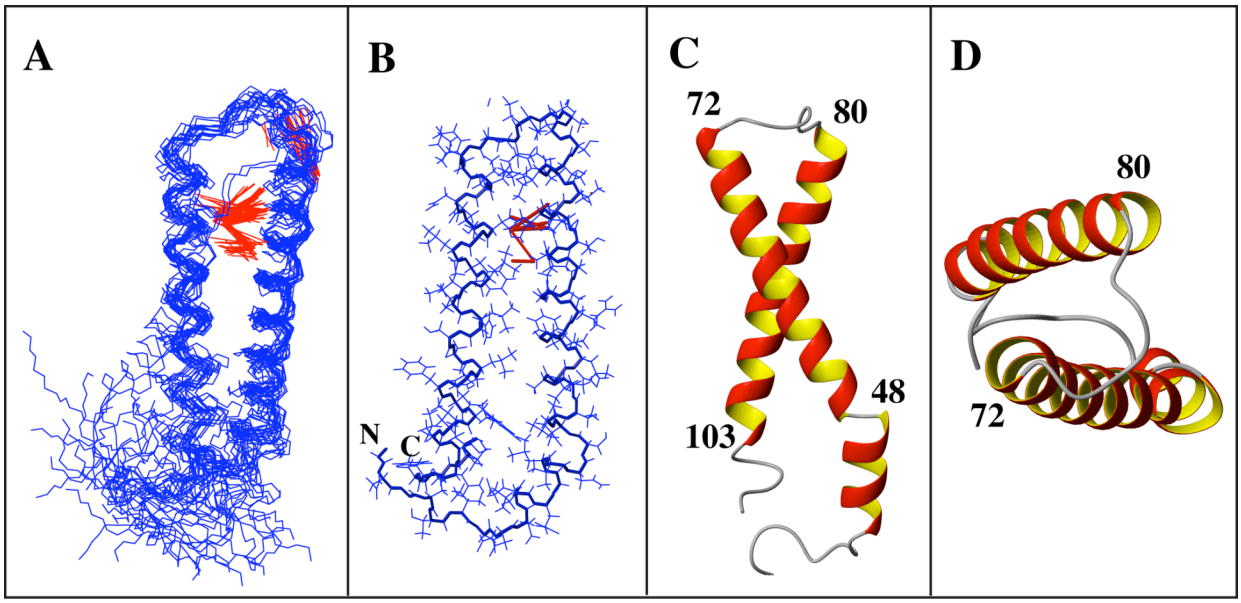


Figure 7: A) Backbone representation of the ensemble of the twenty lowest-energy conformers of Ste2p(G31-T110) superimposed over backbone atoms in the region comprising residues 39-103. Observed long-range NOE contacts are highlighted in red. B) A single conformer from the ensemble additionally displaying the sidechains. C) Structure of a single conformer - view from the side of membrane interior. D) Same as C but viewed from the cytoplasmic side.

When individually superimposing backbone atoms from the TM1 and TM2 helices the RMSD values are 0.40 ± 0.13 Å and 0.57 ± 0.19 Å for backbone atoms of residues 49-72 and 80-103, respectively, while the RMSD is 2.36 ± 0.97 Å when superimposing backbone atoms of residues 49 to 103. For the amphiphilic α-helix of the

N-terminal domain the RMSD value is 0.25 ± 0.10 Å. Hydrogen bonds between carbonyl oxygen atoms of residue i and amide hydrogen atoms of residue i+4 are observed in more than 80% of the structures in the segments 39-47, 49-72 and 80-103. Considering that residues 55-60 display lower H-NOE values (Fig. 5), weaker cross-peaks in the [^{15}N,^{1}H]-HSQC (Fig. 3) and fewer correlations in the corresponding 3D NOESY spectrum (Fig. 6), we suspect that the segment comprising residues 55-60 around the internal Arg58 undergoes slow conformational exchange corresponding to a kink motion. A similar behavior was observed for the residues 80 to 88 indicating that the entire N-terminal half of the 2nd TM helix is destabilized compared to the C-terminal half. The segments encompassing residues 31-38, 71-79 and 104-110 display no contacts characteristic for helices, have decreased H-NOE values and therefore are more flexible. Presently, the number of tertiary contacts between the TM helices remains insufficient to unambiguously establish their relative orientation. However, we could observe unambiguous long-range NOE contacts between methyl groups of the residues Leu66-Val86, Ala63-Val86, Ala63-Leu90 and Val69-Ser75 that help to partially restrain the tertiary structure in the helical regions adjacent to the loop (Fig. 7). No unambiguous long-range contacts involving the p-systems of the aromatic residues were detected.

We have superimposed the experimental structure determined in this work with the corresponding segment from the model of the Ste2p receptor published by Eilers et al.(63) (see Fig. S4). Interestingly, the overall features of the helical hairpin are very similar in both structures (1.8 Å for backbone atoms of TM1 and TM2). TM2 in the experimental structure is slightly rotated counterclockwise along the helix axis. The angle between the two helices is larger in the experimental structure, and the helices are slightly more closely packed in the structure derived from homology modelling.

Amide proton exchange

To investigate to which extent amide protons are protected from solvent exchange we measured reductions in amide proton intensities due to saturation transfer in an [^{15}N, ^{1}H]-HSQC spectrum (Fig S3). Significantly reduced saturation transfer (>80% remaining peak intensity) was observed in the segments 37-48, 52-64, 66-72, 80-107, indicating that

amide moieties from the transmembrane helices are mostly shielded from water access, while more rapid exchange was measured in the N-terminal region (res. 32-36) of the peptide and in the loop Thr72 to Thr78. The observation that residues 37-48 from the N terminus of the protein were protected from solvent exchange is in agreement with the occurrence of a nascent, probably surface-associated, helix in that part of the peptide chain. The data obtained from saturation transfer experiment are consistent with the data obtained from the heteronuclear H-NOE, indicating that regions of increased backbone rigidity and solvent protection extend a couple of residues beyond locations of the putative transmembrane helices.

5.3. Discussion

Despite intense efforts no high-resolution NMR structure of an entire GPCR has been reported to date. Very recently, the nearly complete backbone assignment of sensory rhodopsin in DHPC was published (15). Moreover, although many biophysical studies on peptide surrogates representing regions of IMPs have been conducted, only a few detailed NMR studies on membrane-spanning peptide fragments of GPCRs in detergent micelles appear in the literature. We recently described a solution-state NMR analysis on a fragment comprising the 7^{th} TM domain plus 40 residues from the cytosolic region of Ste2p in DPC micelles (34). A peptide corresponding to the sixth transmembrane domain of Ste2p was also analyzed in lipid bilayers by solid state NMR, and was found to have a very similar structure to that observed in TFE/water (35).

Although important insights into the biophysical properties of peptides corresponding to single TM regions of IMPs have been derived it remains uncertain whether such peptides are good surrogates to learn about the actual structure of these regions in the entire GPCR. Despite the fact that peptides corresponding to single TMs of GPCRs (64) did assume helices in the presence of detergents and in organic aqueous media (27, 65) long-range stabilizing interactions between the individual TM helices may be required for the GPCR fragment to fold into the biologically relevant conformation.

Such long-range interactions might be missing in a single transmembrane fragment of the heptahelical GPCR. Our goal in this investigation was to extend the high-resolution analysis on membrane protein polypeptides to a region of a model GPCR that consisted of two TMs and the intervening loop. Previous studies on the MerF protein, the human Gly transporter and on subunit C of the F_1F_0 ATP synthase provided precedents that peptides, including those corresponding to IMP fragments consisting of two TMs, can fold into a defined tertiary structure in both detergent micelles (66) and organic solvents (41, 67). The only previous NMR study on a double TM fragment of a GPCR was conducted in DMSO on the TM1-TM2 fragment of the cannabinoid receptor (39). This study resulted in nearly complete assignments of the backbone atoms and concluded that this region of the receptor was highly helical, but lacked insight into the tertiary structure of the fragment. Most recently, a detailed NMR analysis of the TM3-TM4 hairpin of the CFTR receptor in perfluorooctanoate micelles led to nearly complete assignments of the backbone resonances and, with the help of a number of specific mutations, provided insights into the structure of the helical hairpin (42).

Here, using a variety of isotope labeling patterns and NMR experiments, we report the nearly complete assignment of the backbone and side chain nuclei in detergent micelles for an 80-residue peptide corresponding to the first two transmembrane domains of Ste2p, the yeast α-factor GPCR. During the course of this project we noted that the use of deuterated LPPG significantly improved spectral quality (see Supp Mat Fig. S2) by both eliminating interfering micellar resonances and decreasing intermolecular relaxation pathways that likely resulted in broadening of peptide resonances. In addition, only when peptides specifically labeled at certain methyl groups (Fig. 4) and NMR experiments suggested by the Kay laboratory (68) were used, could individual methyl resonances be resolved. The combination of deuterated LPPG and methyl-labeled peptides allowed us to identify many medium range NOEs (Fig 6) that defined secondary structure in the TM1 and TM2 helices and, most importantly, to discern long-range NOEs between residues on TM1 and TM2. These latter interactions provide support for the conclusion that in LPPG this fragment folds to a helical hairpin-like structure (Fig. 7).

In addition to defining the secondary structure of the transmembrane regions and identifying the helical hairpin, our investigation revealed the presence of a helix in the

amino-terminal portion of TM1-TM2 that according to its amphiphilic nature is most likely surface-associated. The identification of a helical element in the amine terminal tail of TM1-TM2 may be a biologically significant finding as the N-terminus of GPCRs plays essential roles in their biology. Although studies on chimeric Ste2p from *S. cerevisiae and S. kluyveri* provided evidence that the first 45 residues in the N-terminus of Ste2p from *S. cerevisiae* were not critical for ligand binding specificity or signaling (69), substitutions in residues 47-49 affected binding specificity but not signaling (70). Moreover, deletion of the first 30 residues of the 51 residue N-terminus of Ste2p led to MATα cells that signaled but could not mate (71). The helical region (residues 39-47) of the N-terminus of TM1-TM2 that our NMR studies discerned is likely an amphiphilic helix that interacts with the phospholipid head groups of the LPPG micelles. If similar interactions occur in the native environment, these may help to define the biologically active structure of the pheromone receptor. Previous crystal structures published on GPCRs either do not define their N-termini or reveal some tendency to β-sheet conformations in the termini of rhodopsin family receptors (72). A NMR study on a synthetic peptide corresponding to the first 110 residues of the N-Y4 receptor also found a short helical structure in the extracellular region of this N-terminal GPCR fragment (73, 74). Thus, it is conceivable that the extracellular N-terminal domain in GPCRs contain specific secondary structures or incipient structures that can be stabilized on interaction with either ligands or other accessory proteins involved in signal transduction pathways.

The conformation of the segments corresponding to the putative TM helices 1 and 2 had reasonably low RMSD values when superimposing the individual helices. The α-helical character is very well supported by numerous medium-range NOEs. The observation of these medium range NOEs indicate that the helices in the 2-TM construct are much more stable than the single TM helix we previously described from the 7^{th} TM of the same receptor in DPC micelles (34). The TM1 helix is destabilized in the G56-G60 region, and the helices cross each other and splay apart near the putative extracellular surface of this GPCR domain. At least four independent contacts between residues adjacent to the first intracellular loop were established. No such contacts were observed between residues near the center or the extracellular face of the helical hairpin. This finding could reflect the fact that contacts with additional TMs are necessary to stabilize

the structure of this region of Ste2p. Although the spatial relationship between the two TMs is presently underdetermined it is important that the polypeptide appears to take a turn in the micellar environment rather than assuming a large distribution of structures. At present we are evaluating experimental approaches to gain additional restraints that help to more accurately define the relative orientation of TM1 and TM2.

There are several aspects of the NMR structure of TM1-TM2 that should be considered in the context of biological and biochemical information about the function of this peptide. Previous analyses of Ste2p biology show that the first extracellular loop and the extracellular end of TM1 are involved in both the binding of pheromone and the signal transduction pathway (69, 75-77). Crosslinking studies have revealed that Tyr^{13} of α-factor contacts residues 55-59 of TM1 (78), and we recently determined that the contact points are R58 and Cys59 by photochemical and oxidative crosslinking approaches (Becker and Naider, submitted for publication). Thus, it seems clear that the carboxyl terminus of the α-factor must penetrate into the TM interior of Ste2p upon binding to this receptor. If this is correct it would not be possible for TM1 to be tightly packed against TM2 and our finding of destabilization in the G56-G60 region and splaying apart of the two TMs would be consistent with the crosslinking results. Part of the driving force for the destabilization most likely stems from the occurrence of polar and, in particular, charged residues within the membrane. For example, transferring an Arg residue into the membrane interior requires 1.8 kcal mol^{-1} (79). Even if the charges from these residues are partially compensated by polar residues placed in other TM helices they may still introduce some conformational instability that may be important for the creation of the pheromone binding pocket.

In conclusion we report here an NMR structure for Ste2p(G31-T110) of the GPCR mating receptor from *S. cerevisiae* in LPPG micelles. Almost complete resonance assignments were accomplished for the 80-residue fragment, representing more than 25% of the residues from the core of this receptor. The conformation was determined without introducing any artificial restraints and its secondary structure is well–defined. A few interhelical contacts demonstrate that the protein is folded in micelles into a helical hairpin that splays apart at the termini. A region of the receptor predicted to be in the N-terminal receptor tail formed a helix that likely interacts with the surface of the micelles.

To our knowledge this structure is the first reported for a double TM containing fragment of a GPCR in lipid. Its tendency to assume a specific tertiary structure supports the use of GPCR fragments as models to discern the structure of the intact receptor.

5.4. Materials and Methods

Materials

Deuterated water, α-ketobutyric [$^{13}C_4$, 98%; 3,3-D_2, 98%] and α-ketoisovaleric [1,2,3,4,-$^{13}C_4$, 99%; 3,4,4,4-D_4, 98%] acids were bought from CAMBRIDGE ISOTOPES and (deuterated) lipids were purchased from AVANTI POLAR LIPIDS (USA). All other chemicals used were ordered from SIGMA-ALDRICH or FLUKA.

Cloning, expression and purification of isotopically-labeled Ste2p(G31-T110)[TM1-TM2]

The cloning, expression, and isolation of Ste2p(G31-T110) labelled with ^{15}N, ^{13}C and ^{2}H; [^{15}N,^{13}C,^{2}H]-TM1-TM2 (G^{31}NGST ITFDE L^{41}QGLV NSTVT Q^{51}AILF GVRSG A^{61}AALT LIVVW I^{71}TSRS RKTPI F^{81}IINQ VSLFL I^{91}ILHS ALYFK Y^{101}LLSN YSSVT) were carried out using procedures described previously (44). This peptide contains 4 replacements of natural residues (3 methionines were replaced with leucine, valine and isoleucine and Cys was replaced with Ser) to enable the CNBr cleavage and to stabilize the peptide against oxidation. The logic of the exact replacements used was discussed previously (44). Cys59 can be replaced by Ser without any effect on the biological activity of Ste2p (45) and replacement of individual methionine residues with leucine, valine or isoleucine was biologically acceptable (46). TM1-TM2 peptides selectively labelled with [^{15}N]-alanine, [^{15}N]-isoleucine, [^{15}N]-leucine, [^{15}N]-valine or [^{15}N]-phenylalanine were prepared in defined minimal medium supplemented with all

unlabelled amino acids and an excess of the [^{15}N]-labelled amino acid as described by Cohen et al. (44).

A sample of TM1-TM2 that contained protonated methyl groups in an almost fully deuterated background was prepared as described by Tugarinov et al. (47). Briefly, BL21-AI cells containing pLC01 were streaked onto LBAmp plates and incubated overnight at 37°C. A 6 ml LBAmp culture was inoculated with one colony from the overnight growth and was incubated at 37°C, 250 rpm to OD_{600} 0.7-0.8. These cells were pelleted and resuspended in M9 minimal media in H_2O to OD_{600} ~0.05-0.1and then incubated at 37°C as above to OD_{600} 0.6. The cells were once again pelleted and resuspended in 100 ml M9 minimal media in D_2O containing $^{13}C/^{2}H$-glucose and $^{15}NH_4Cl$ (M9/D_2O) and cells were grown until an OD_{600} ~0.4-0.5 was reached. The cells were then diluted to 200 ml with M9/D_2O, incubated as above to OD_{600} ~0.4-0.5 and then diluted to 1L in M9 medium in D_2O supplemented with 70 mg/L α-ketobutyric acid ($^{13}C_4$, 98%; 3,3-D_2, 98%) and 120 mg/L α-ketoisovaleric acid (1,2,3,4-$^{13}C_4$, 99%; 3,4',4',4'-D_4, 98%). These cells were incubated at 37°C, 250 rpm to OD_{600} ~0.3-0.4 and expression was induced with 0.5% L-arabinose, grown at 37°C, 250 rpm for 6 to 8 hours and harvested by centrifugation. Further details of purification are described in Cohen et al. (44).

NMR sample preparation

Partially, uniformly or selectively $^{15}N/^{13}C/^{2}H/^{1}H$-labeled TM1-TM2 NMR samples were obtained by dissolving the peptide (0.2-1 mg) and detergent LPPG (28.4 mg) in sodium phosphate buffer (200µl, 20mM, pH = 6.4, 1-2 min of shaking), followed by sonication (2 x 15 min) and incubation (30 min) at 37°C prior to transferring to a Shigemi NMR tube. The final concentrations used for NMR measurements were ~0.1-0.4mM and 200mM for the peptide and LPPG, respectively. Following this procedure we were able to obtain good quality and reproducible [^{15}N, ^{1}H]-HSQC spectra with intense signals, indicating proper incorporation of the polypeptide into the micelle. The samples prepared using this method were sufficiently stable for measurement of NMR spectra at 320K, and displayed degradation in the form of a white precipitate only after 2 weeks in

case of fully protonated peptide samples and over 4 weeks in case of partially deuterated peptide samples. Two dimensional, 3D triple-resonance and ^{15}N-resolved NOESY spectra were recorded at 320K using a Bruker AV700 spectrometer equipped with a triple-resonance cryoprobe. The ^{13}C-edited HSQC and NOESY spectra centered on methyl (19ppm), aliphatic (39 ppm) and aromatic (125 ppm) carbons were recorded at 320K using a 900 MHz spectrometers at the New York Center for Structural Biology and at the Swiss Federal Institute of Technology. All proton chemical shifts were referenced to the water line at 4.48 ppm at 47°C, from which the nitrogen and carbon scales were derived indirectly by using the conversion factors of 0.10132900 (^{15}N) and 0.25144954 (^{13}C). Chemical shifts were deposited in the BMRB database under the accession code 15995.

NMR spectroscopy

Sequence-specific resonance assignment was accomplished based on a set of triple-resonance experiments as well as ^{15}N- and ^{13}C- resolved NOESY spectra. Backbone assignment was performed based on the set of HNCO/HNCACO experiments (48) [2048(^{1}H)*40(^{15}N)*128(^{13}C) complex data points; $t_3$max 105ms, $t_2$max 14.1, $t_1$max 16.5ms] and HNCA/HNCOCA experiments (48) [2048(^{1}H)*40(^{15}N)*128(^{13}C) complex data points; $t_3$max 105ms, $t_2$max 14.1, $t_1$max 4.9ms]. HNCACB and CBCA(CO)NH spectra (49) [2048(^{1}H)*40(^{15}N)*128(^{13}C) complex data points; $t_3$max 105ms, $t_2$max 14.1ms, $t_1$max 4.9ms] were evaluated to additionally support assignments made and to derive information on Cβ chemical shifts. Sidechain resonance assignment started with hCCH-TOCSY experiment (50, 51) [2048(^{1}H)*50(^{13}C)*100(^{13}C) complex points, $t_3$max 105ms, $t_2$max 4.7ms, $t_1$max 9.5ms] using a B_1 field of 8.3 kHz for the TOCSY spin-lock. Finally chemical shifts were obtained by picking peaks in a 13.3ms constant-time (ct)-[^{13}C, ^{1}H]-HSQC spectrum (52). Unfortunately signal-to-noise in the hCCH data set was insufficient so that extensive use of ^{13}C-resolved NOESYs had to be made. Methyl groups of Val(Hγ), Ile(Hδ1) and Leu(Hδ1/δ2) residues were assigned using a HMCMCBCANH experiment developed by Kay (43) [2048(^{1}H)*40(^{15}N)*36(^{13}C) complex data points; $t_3$max 105ms, $t_2$max 12.8ms, $t_1$max 6.8ms for Val and Ile residues

and 2048*40(^{15}N)*60(^{13}C) complex data points; $t_3$max 105ms, $t_2$max 12.8ms, $t_1$max 9.5ms for Leu residues]. Peak positions were adjusted to the 13.3ms ct-[^{13}C, ^{1}H]-HSQC spectrum and cross-validated against the ^{13}C-resolved NOESY experiments recorded on fully-protonated and partially-methyl-protonated samples. The aromatic ring systems of Tyr, His and Trp residues were picked in a 8.8ms ct-[^{13}C, ^{1}H]-HSQC and correlated with β-carbons via the HBCBCGCDHD experiment (53) [2048(^{1}H)*58(^{13}C) complex points, $t_2$max 91ms, $t_1$max 4.1ms] whenever possible or via NOEs from the ^{13}C-NOESY centered on aromatic carbons. All chemical shifts were finally derived from peaks in the [^{15}N, ^{1}H]-HSQC and the ct-[^{13}C, ^{1}H]-HSQC spectra. All experiments employed pulsed-field gradients (54).

The extent of amide hydrogen exchange was probed by recording a [^{15}N, ^{1}H]-HSQC experiment, in which low-power irradiation was applied on the water resonance during the relaxation delay. Peak volumes were determined in this experiment and their values relative to a reference experiment conducted in the absence of irradiation but with otherwise identical parameters were computed (see Fig. S3). ^{15}N-relaxation data were recorded using proton-detected version of the ^{15}N{^{1}H} steady-state NOE experiment (55, 56) using a 2.7s recycle delay.

Data were usually extended by a factor of two using linear-prediction in the indirect dimensions and processed within the Bruker spectrometer software TOPSPIN 2.0. Processed data were transferred into the program CARA (57) for data analysis. Integration of peak volumes was performed with the SPSCAN/XEASY (58) or CARA software. MOLMOL was used to calculate RMSD values, and to prepare structural representations (59).

Structure calculation

Distance restraints were obtained from ^{15}N-resolved NOESY spectra recorded on [^{15}N,^{13}C]- and [^{15}N,^{2}H]- labeled TM1-TM2 samples with mixing times of 70 and 200ms, respectively, and from ^{13}C-resolved NOESY spectra recorded on [^{15}N,^{13}C]- and [^{15}N,^{13}C,^{2}H(^{1}H(methyl) - Ile, Leu, Val)]- labeled TM1-TM2 samples with mixing times of 100 and 200 ms respectively. In addition, dihedral angle restraints derived from

TALOS (60) using chemical shifts of Hα, Cα, Cβ, C' and ^{15}N nuclei were added. Structures were calculated using the standard simulated-annealing protocol for molecular dynamics in torsion angle space as implemented in the program CYANA (61). The final CYANA calculation was performed with 100 randomized starting structures, and the 20 CYANA conformers with the lowest target function values were selected to represent the NMR ensemble. The conformers were analyzed, and figures were prepared within the program MOLMOL (59). The coordinates have been deposited in the PDB database under accession code 2k9p.

Acknowledgements

This work was supported by research grants GM22086 (FN) and GM22087 (JMB) from the National Institutes of Health and a grant from the Alfred Werner Legat (OZ). Prof. Fred Naider is a member of the New York Structural Biology Center. The Center is a STAR center supported by the New York State Office of Science, Technology, and Academic Research. The 900 MHz spectrometer at the NYSBC was purchased with funds from NIH, USA, the Keck Foundation, New York State, and the NYC Economic Development Corporation.

5.5. References

1. Boyd, D., Schierle, C. and Beckwith, J. (1998) *Protein Sci.* **7**, 201-205
2. Stevens, T. J. and Arkin, I. T. (2000) *Proteins* **39**, 417-420
3. Hopkins, A. L. and Groom, C. R. (2002) *Nat. Rev. Drug Discov.* **1**, 727-730
4. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. and Miyano, M. (2000) *Science* **289**, 739-745
5. Rasmussen, S. G. F., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., Schertler, G. F. X., Weis, W. I. and Kobilka, B. K. (2007) *Nature* **450**, 383-U4
6. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G. and Schertler, G. F. (2008) *Nature* **454**, 486-491
7. Sarramegn, V., Muller, I., Milon, A. and Talmont, F. (2006) *Cell. Mol. Life Sci.* **63**, 1149-1164
8. Grisshammer, R., White, J. F., Trinh, L. B. and Shiloach, J. (2005) *J. Struct. Funct. Genomics* **6**, 159-163
9. Page, R. C., Moore, J. D., Nguyen, H. B., Sharma, M., Chase, R., Gao, F. P., Mobley, C. K., Sanders, C. R., Ma, L., Sönnichsen, F. D., Lee, S., Howell, S. C., Opella, S. J. and Cross, T. A. (2006) *J. Struct. Funct. Genom.* **7**, 51-64
10. Sanders, C. R. and Sönnichsen, F. (2006) *Magn. Reson. Chem.* **44**, S24-40
11. Hu, J., Qin, H., Li, C., Sharma, M., Cross, T. A. and Gao, F. P. (2007) *Prot. Sci.* **16**, 2153-2165
12. Williamson, P. T., Verhoeven, A., Miller, K. W., Meier, B. H. and Watts, A. (2007) *Proc. Natl. Acad. Sci. U S A* **104**, 18031-18036
13. Vosegaard, T., Kamihira-Ishijima, M., Watts, A. and Nielsen, N. C. (2008) *Biophys. J.* **94**, 241-250
14. Luca, S., White, J. F., Sohal, A. K., Filippov, D. V., van, B., JH, Grisshammer, R. and Baldus, M. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 10706-10711
15. Gautier, A., Kirkpatrick, J. P. and Nietlispach, D. (2008) *Angew. Chem. Int. Ed. Engl.* **47**, 7297-7300
16. Popot, J. L. and Engelman, D. M. (2000) *Annu. Rev. Biochem.* **69**, 881-922
17. White, S. H. and Wimley, W. C. (1999) *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319-365
18. Senes, A., Engel, D. E. and DeGrado, W. F. (2004) *Curr. Opin. Struct. Biol.* **14**, 465-479

19. Barsukov, I. L., Abdulaeva, G. V., Arseniev, A. S. and Bystrov, V. F. (1990) *Eur. J. Biochem.* **192**, 321-327

20. Grabchuk, I. A., Orekhov, V. Y. and Arseniev, A. S. (1996) *Pharm. Acta Helv.* **71**, 97-102

21. Lomize, A. L., Pervushin, K. V. and Arseniev, A. S. (1992) *J. Biomol. NMR* **2**, 361-372

22. Pervushin, K. V. and Arseniev, A. S. (1992) *FEBS Lett.* **308**, 190-196

23. Pervushin, K. V., Arseniev, A. S., Kozhich, A. T. and Ivanov, V. T. (1991) *J. Biomol. NMR* **1**, 313-322

24. Pervushin, K. V., Orekhov, V. Y., Popov, A. I., Musina, L. Y. and Arseniev, A. S. (1994) *Eur. J. Biochem.* **219**, 571-83.

25. Chopra, A., Yeagle, P. L., Alderfer, J. A. and Albert, A. D. (2000) *Biochim. Biophys. Acta* **1463**, 1-5

26. Katragadda, M., Chopra, A., Bennett, M., Alderfer, J. L., Yeagle, P. L. and Albert, A. D. (2001) *J. Pept. Res.* **58**, 79-89

27. Arshava, B., Taran, I., Xie, H., Becker, J. M. and Naider, F. (2002) *Biopolymers* **64**, 161-176

28. Xie, H., Ding, F. X., Schreiber, D., Eng, G., Liu, S. F., Arshava, B., Arevalo, E., Becker, J. M. and Naider, F. (2000) *Biochemistry* **39**, 15462-15474

29. Reddy, A. P., Tallon, M. A., Becker, J. M. and Naider, F. (1994) *Biopolymers* **34**, 679-689

30. Arshava, B., Liu, S. F., Jiang, H., Breslav, M., Becker, J. M. and Naider, F. (1998) *Biopolymers* **46**, 343-357

31. Estephan, R., Englander, J., Arshava, B., Samples, K. L., Becker, J. M. and Naider, F. (2005) *Biochemistry* **44**, 11795-11810

32. Naider, F., Arshava, B., Ding, F. X., Arevalo, E. and Becker, J. M. (2001) *Biopolymers* **60**, 334-350

33. Naider, F., Khare, S., Arshava, B., Severino, B., Russo, J. and Becker, J. M. (2005) *Biopolymers* **80**, 199-213

34. Neumoin, A., Arshava, B., Becker, J., Zerbe, O. and Naider, F. (2007) *Biophys. J.* **93**, 467-482

35. Valentine, K. G., Liu, S. F., Marassi, F. M., Veglia, G., Opella, S. J., Ding, F. X., Wang, S. H., Arshava, B., Becker, J. M. and Naider, F. (2001) *Biopolymers* **59**, 243-256

36. Thévenin, D., Roberts, M. F., Lazarova, T. and Robinson, C. R. (2005) *Biochemistry* **44**, 16239-16245

37. Thevenin, D. and Lazarova, T. (2008) *Protein Sci.* **17**, 1188-1199

38. Kerman, A. and Ananthanarayanan, V. S. (2005) *Biochim. Biophys. Acta* **1747**, 133-140

39. Zheng, H., Zhao, J., Sheng, W. and Xie, X. Q. (2006) *Biopolymers* **83**, 46-61

40. Kerman, A. and Ananthanarayanan, V. S. (2007) *Biochim. Biophys. Acta* **1768**, 1199-1210

41. Ma, D., Liu, Z., Li, L., Tang, P. and Xu, Y. (2005) *Biochemistry* **44**, 8790-8800

42. Wehbi, H., Gasmi-Seabrook, G., Choi, M. Y. and Deber, C. M. (2008) *Biochim. Biophys. Acta* **1778**, 79-87

43. Tugarinov, V. and Kay, L. E. (2003) *J. Am. Chem. Soc.* **125**, 13868-13878

44. Cohen, L. S., Arshava, B., Estephan, R., Englander, J., Kim, H., Hauser, M., Zerbe, O., Ceruso, M., Becker, J. M. and Naider, F. (2008) *Biopolymers* **90**, 117-130

45. Hauser, M., Kauffman, S., Lee, B. K., Naider, F. and Becker, J. M. (2007) *J. Biol. Chem.* **282**, 10387-10397

46. Martin, N. P., Celic, A. and Dumont, M. E. (2002) *J. Mol. Biol.* **317**, 765-788

47. Tugarinov, V., Kanelis, V. and Kay, L. E. (2006) *Nature Protocols* **1**, 749-754

48. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. and Kay, L. E. (1994) *J. Am. Chem. Soc.* **116**, 11655-11666

49. Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. and Kay, L. E. (1996) *J. Am. Chem. Soc.* **118**, 6570-6579

50. Kay, L., Xu, G. Y., Singer, A. U., Muhandiram, D. R. and Forman-Kay, J. D. (1993) *J. Magn. Reson. Ser. B* **101**, 333-337

51. Olejniczak, E. T., Xu, R. X. and Fesik, S. W. (1992) *J. Biomol. NMR* **2**, 655-659

52. Vuister, G. W. and Bax, A. (1992) *J. Mag. Reson.* **98**, 428

53. Yamazaki, T., Formankay, J. D. and Kay, L. E. (1993) *J. Am. Chem. Soc.* **115**, 11054-11055

54. Keeler, J., Clowes, R. T., Davis, A. L. and Laue, E. D. (1994) *Methods Enzymol.* **239**, 145-207

55. Noggle, J. H. and Schirmer, R. E. (1971) *The Nuclear Overhauser Effect - Chemical Applications*, Academic Press, New York

56. Palmer, A. G. (2004) *Chem. Rev.* **104**, 3623-3640

57. Keller, R. (2004) *The Computer Aided Resonance Assignment*, CANTINA Verlag, Goldau

58. Bartels, C., Xia, T.-h., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR* **6**, 1-10

59. Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.* **14**, 51–55

60. Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR* **13**, 289-302

61. Güntert, P. (2004) *Methods Mol. Biol.* **278**, 353-378

62. Kabsch, W. and Sander, C. (1983) *Biopolymers* **22**, 2577-2637

63. Eilers, M., Hornak, V., Smith, S. O. and Konopka, J. B. (2005) *Biochemistry* **44**, 8959-8975

64. Hunt, J. F., Rath, P., Rothschild, K. J. and Engelman, D. M. (1997) *Biochemistry* **36**, 15177-15192

65. Katragadda, M., Alderfer, J. L. and Yeagle, P. L. (2001) *Biophys. J.* **81**, 1029-1036

66. Howell, S. C., Mesleh, M. F. and Opella, S. J. (2005) *Biochemistry* **44**, 5196-5206

67. Girvin, M. E., Rastogi, V. K., Abildgaard, F., Markley, J. L. and Fillingame, R. H. (1998) *Biochemistry* **37**, 8817-8824

68. Tugarinov, V. and Kay, L. (2005) *ChemBioChem* **6**, 1567-1577

69. Sen, M. and Marsh, L. (1994) *J. Biol. Chem.* **269**, 968-973

70. Sen, M., Shah, A. and Marsh, L. (1997) *Curr. Genet.* **31**, 235-240

71. Shi, C., Kaminskyj, S., Caldwell, S. and Loewen, M. C. (2007) *Proc. Natl. Acad. Sci. U S A* **104**, 5395-5400

72. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. and Ernst, O. P. (2008) *Nature* **454**, 183-187

73. Zou, C., Kumaran, S., Markovic, S., Walser, R. and Zerbe, O. (2008) *ChemBioChem* **9**, 2276-2284

74. Zou, C., Naider, F. and Zerbe, O. (2008) *J. Biomol. NMR* in press

75. Lee, B. K., Khare, S., Naider, F. and Becker, J. M. (2001) *J. Biol. Chem.* **276**, 37950-37961

76. Akal-Strader, A., Khare, S., Xu, D., Naider, F. and Becker, J. M. (2002) *J. Biol. Chem.* **277**, 30581-30590

77. Lin, J. C., Duell, K. and Konopka, J. B. (2004) *Mol. Cell. Biol.* **24**, 2041-2051

78. Son, C. D., Sargsyan, H., Naider, F. and Becker, J. M. (2004) *Biochemistry* **43**, 13193-13203

79. White, S. H. and Wimley, W. C. (1998) *Biochim. Biophys. Acta* **1376**, 339-52.

5.6. Supplementary material

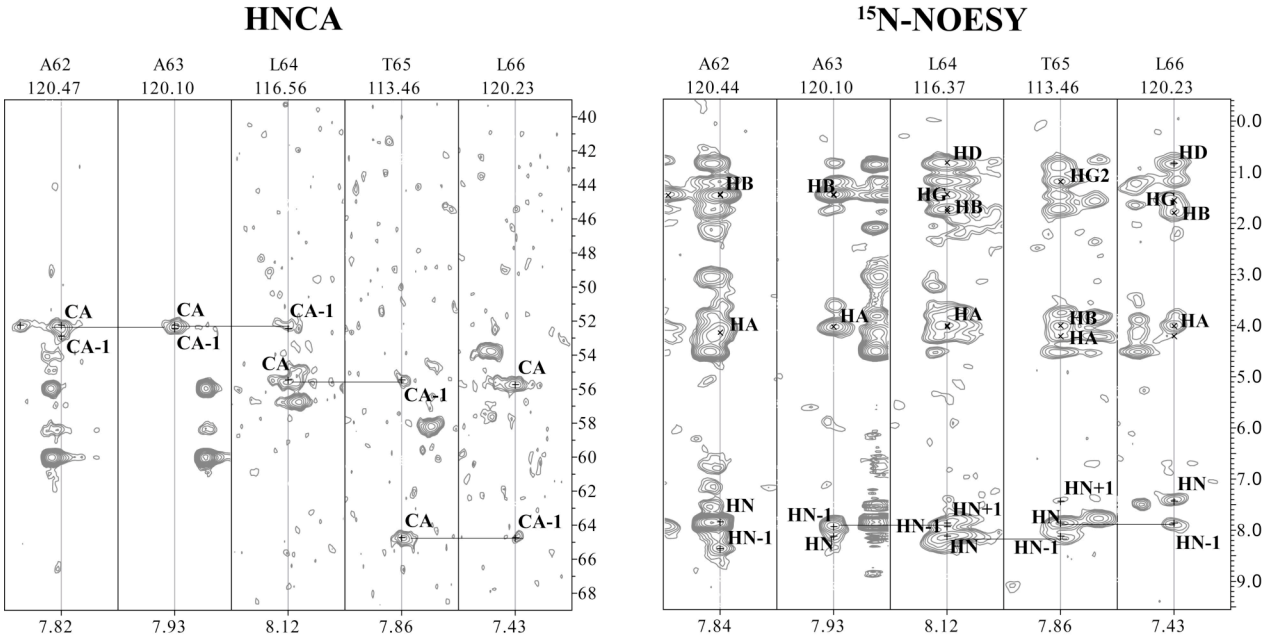


Figure S1. Strips from the 3D HNCA (left) and the ^{15}N-resolved NOESY (right) spectra extracted at various amide proton positions displaying the assignment and validation processes of the ^{15}N, ^{13}C and ^{1}H chemical shifts. The ^{15}N and ^{1}H chemical shifts, at which the strips were extracted, are displayed above and below the strips, respectively.

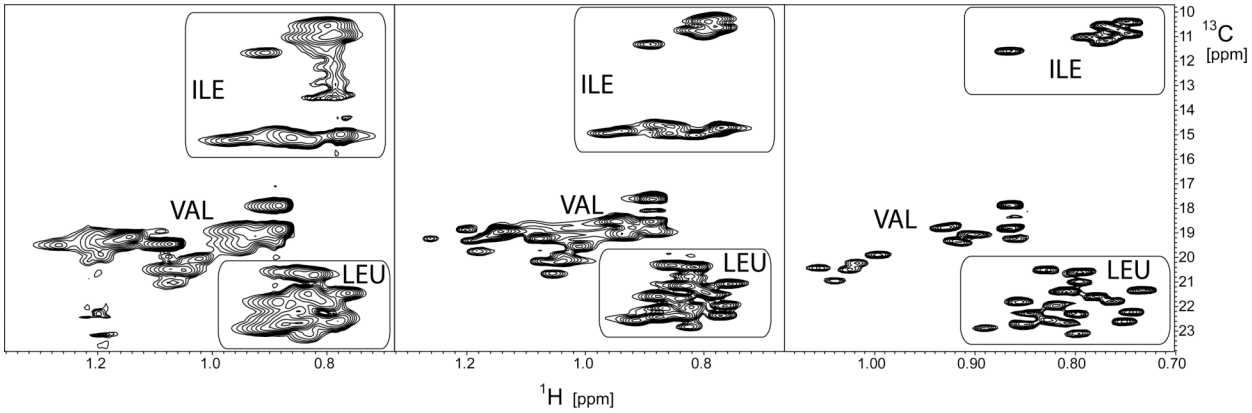


Figure S2. Methyl region from the ct-[^{13}C, ^{1}H]-HSQC of the samples: Left – [^{13}C,^{15}N]-labelled TM1-TM2 sample in the LPPG solution, recorded on a 700MHz magnet; Middle – [^{13}C,^{15}N]-labelled TM1-TM2 sample in the d_{36}-LPPG solution, recorded on a 900MHz magnet; Right - [^{13}C,^{15}N,^{2}H,^{1}H(Methyl – Ile, Leu, Val)]-labelled TM1-TM2 sample in d_{36}-LPPG solution, recorded on a 900MHz magnet. The regions containing methyl groups of Ile or Leu residues are marked with boxes.

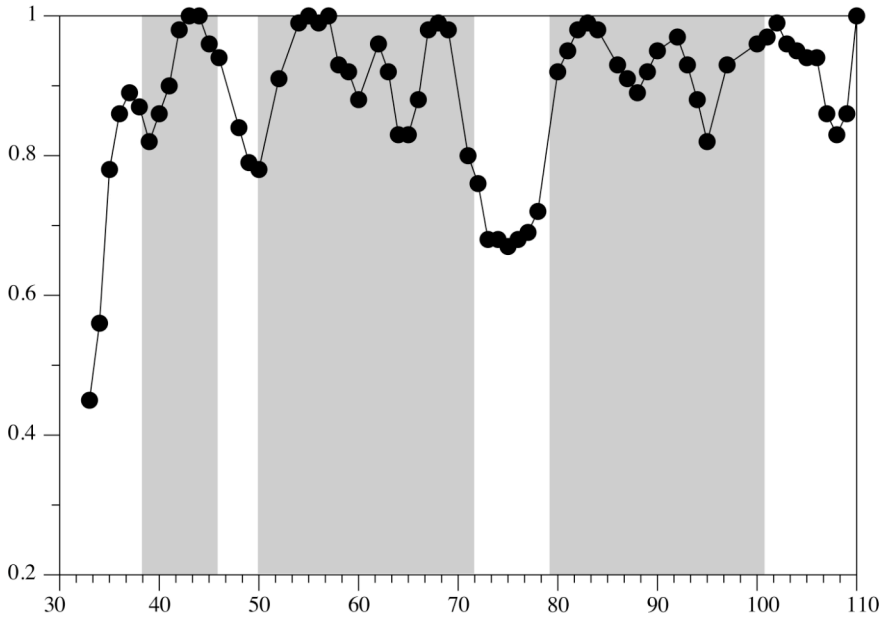


Figure S3. Relative peak volumes of signals computed from [^{15}N,^{1}H]-HSQC spectra recorded in the presence of low-power presaturation on the water resonance during the relaxation delay relative to a reference experiment without presaturation. Highlighted with grey are the predicted α-helical regions of TM1-TM2 peptide as well as the additionally observed N-terminal amphiphilic helix.

Table S1. Information on the structure calculation.

Distance restraints	Total	1247
	Intra-residual	439
	Sequential (i - j = 1)	378
	Medium (i - j = 2, 3, 4)	406
	Long-range	24
Dihedral angle restraints	Total	753
RMSD (Å)		
	Asp39-Ser47 backbone	0.25 ± 0.10
	Asp39-Ser47 all heavy atoms	1.32 ± 0.26
	Val49-Thr72 backbone	0.40 ± 0.13
	Val49-Thr72 all heavy atoms	1.20 ± 0.31
	Ile80-Leu103 backbone	0.57 ± 0.19
	Ile80-Leu103 all heavy atoms	1.38 ± 0.24
	Asp39-Leu103 backbone	2.36 ± 0.97
	Asp39-Leu103 all heavy atoms	3.28 ± 1.03
Structure check (Average %)		
according to Ramachandran plot	Gly31-Thr110 most favored	83.9
	Gly31-Thr110 additionally allowed	15.2
	Gly31-Thr110 generously allowed	0.9
	Gly31-Thr110 disallowed	0
NOE constraint violations	Number > 0.1 Å	7
	Maximum (Å)	0.53
Dihedral angle constraint violations	Number > 2.5 degrees	0
AMBER energies (kcal/mol)	Total	-1029.1
	Van der Waals	284.5
	Electrostatic	-1916.5

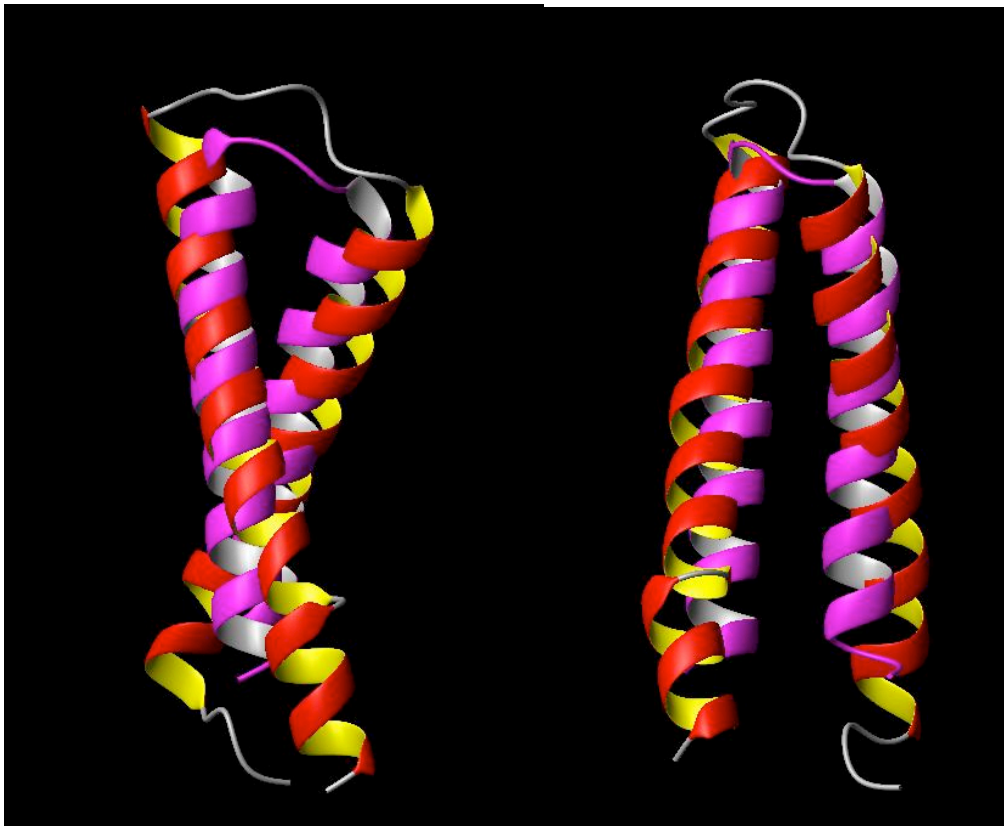


Figure S4. Superposition of the lowest-energy conformer from the calculated 20-conformers bundle with the structure derived from homology modeling (Eilers *et al.*) fitted for backbone atoms of residues 49-72 and 80-103. Right and left representations differ by a 90^{o} rotation about the bilayer normal. The red/yellow representation corresponds to the experimental structure and the pink/gray to the modeled structure.

Publications

1. Mares J., Müller J., Skirgailiene A., Neumoin A., Bewley C., Schmidt R., Zerbe O. (2006) Chembiochem. 7(11): 1764-73.

2. Zerbe, O., Neumoin, A., Mares, J., Walser, R., Zou, C. (2006). J. Recept. Signal Transduct. Res. 26(5-6): 487-504.

3. Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R., Zerbe, O. (2007). J. Am. Chem. Soc. 129(28): 8811-7

4. Neumoin, A., Arshava, B., Becker, J., Zerbe, O., Naider, F. (2007). Biophys. J. 93(2): 467-82

5. Neumoin, A., Cohen, L., Arshava, B., Tantry, S., Becker, J., Zerbe, O., Naider, F. *submitted for publication*.

6. Mares, J., Walser, R., Neumoin, A., Zou, C., Zerbe, O. *manuscript in preparation*.

Alexey V. Neumoin

Date of birth: 14.01.1981
Place of birth: Sarov, Russian Federation
Citizenship: Russian Federation
Marital status: Single

Address for correspondence
Prof. Dr. O. Zerbe, NMR group
Organic Chemistry Department
University of Zurich
Winterthurerstr. 190, CH-8057
Zurich, Switzerland

Permanent address
Bergstrasse 128, CH-8032
Zurich, Switzerland
Tel: +41 44 635-42-73
Mobile: +41 76 369-69-16
Fax: +41 44 635-68-82

E-mail: neumoin@oci.uzh.ch

EDUCATION

02/2004-12/2008	NMR Group, Organic Chemistry Institute, University of Zurich, Switzerland PhD in biophysical chemistry Structural investigation of GPCR fragments and ligands in water and membrane-mimicking environment by NMR spectroscopy
09/1998-06/2003	Enzymology Division, Chemistry Department Moscow State University, Russian Federation 19/06/2003 Diploma with honor in Chemical Enzymology/Biotechnology
09/1994-06/1998	Lyceum of Natural Sciences, Vologda, Russian Federation Form with a chemistry bias
09/1987-06/1994	Middle School #1, Vologda, Russian Federation Form with English language bias

SCIENTIFIC PUBLICATIONS

1. Neumoin A, Mares J, Lerch-Bader M, Bader R, Zerbe O., J Am Chem Soc. **2007**, 129(28), 8811-7.
2. Neumoin A, Arshava B, Becker J, Zerbe O, Naider F., Biophys J. **2007**, 93(2), 467-82.
3. Zerbe O, Neumoin A, Mares J, Walser R, Walser R, Zou C., J Recept Signal Transduct Res. **2006**, 26(5-6), 487-504.
4. Mares J, Muller JU, Skirgailiene A, Neumoin A, Bewley CA, Schmidt RR, Zerbe O., Chembiochem. **2006**, 7(11), 1764-73.
5. Neumoin, A., Cohen, L., Arshava, B., Tantry, S., Becker, J., Zerbe, O., Naider, F. **2009**, *submitted for publication*.
6. Mares, J., Walser, R., Neumoin, A., Zou, C., Zerbe, O. *manuscript in preparation*.

INTERNATIONAL COLLABORATION

06/2002-08/2002	3 months research visiting to Heraklion (Greece), Grant of European Community INTAS Ref. No: 00-00870
07/2001	2 weeks visiting to China, Grant of RFFI-GFEN # 99-03-39064.

OUTSIDE WORKING EXPERIENCE

08/2003-01/2004	Client Support Department of a cable TV Divo-TV (Moscow). Full telephone client support.
12/2002-05/2003	Client Support Department of an electronic payment group E-port (Moscow). Full telephone client support

LANGUAGES

English – *fluent,* Russian – *native,* German – good (Goethe certificate B2 – "gut")

COMPUTER SKILLS:

MAC OSX, SUSE Linux OS and software, OS Windows – advanced user, minor administration experience; NMR recording and processing software (Bruker), MS Office, Adobe CS – experienced user; proficient in use of Internet resources and online databases.

OUTSIDE INTERESTS

Active sports, travelling, music, computers and information technologies

REFERENCE

Prof. Dr. Oliver Zerbe
NMR Group, Organic Chemistry Institute, University of Zurich
Winterthurerstr. 190, CH-8057 Zurich, Switzerland
Tel: +41 44 635-42-63, Fax: +41 44 635-68-82
E-mail: oliver.zerbe@oci.uzh.ch